

UNIVERSITÉ DE SHERBROOKE

CARACTÉRISATION STRUCTURALE ET BIOPHYSIQUE DU COMPLEXE C-MYC/MID2 – ÉLUCIDATION D'UNE PLATFORME MOLÉCULAIRE POUR LA DÉCOUVERTE DE NOUVEAUX INHIBITEURS DE C-MYC

Par

JEAN-MICHEL MOREAU

Programme de Biochimie

Mémoire présenté à la Faculté de Médecine et de Sciences de la Santé

En vue de l'obtention du grade de
Maître ès Sciences (M.Sc.) en Biochimie

Sherbrooke, Québec, Canada

Juillet 2019

Membres du jury d'évaluation

Pr. Pierre Lavigne, département de Biochimie

Pr. Xavier Roucou, département de Biochimie

Pr. Éric Marsault, département de Pharmacologie

SOMMAIRE

CARACTÉRISATION STRUCTURALE ET BIOPHYSIQUE DU COMPLEXE C-MYC/MID2 – ÉLUCIDATION D'UNE PLATFORME MOLÉCULAIRE POUR LA DÉCOUVERTE DE NOUVEAUX INHIBITEURS DE C-MYC

Par
Jean-Michel Moreau
Programme de Biochimie

Mémoire présenté à la Faculté de Médecine et des Sciences de la Santé
En vue de l'obtention du grade de Maître ès Sciences (M.Sc.) en Biochimie
Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

c-Myc est un facteur de transcription de type bHLH-LZ qui joue un rôle central dans la croissance et la prolifération cellulaire en contrôlant l'expression d'une pléthore de gènes impliqués dans la ribogénèse, le métabolisme, l'apoptose et le cycle cellulaire. Les niveaux d'expression de c-Myc sont dérégulés et surélevés suite à l'activation d'oncogène pilotes ou à la translocation de son gène. Elle devient alors oncogénique et contribue à la tumorigénèse d'un vaste éventail de cancer. De plus, les cellules tumorales surexprimant c-Myc deviennent dépendantes à son activité transcriptionnelle exacerbée et cette dépendance représente le talon d'Achille des cellules tumorales. En effet, son inhibition avec une protéine dominante négative (Omomyc) dans plusieurs modèles cellulaires et murins de cancer mène à un effondrement des tumeurs établies. Grâce à ces travaux, c-Myc est maintenant reconnue comme une cible thérapeutique importante et beaucoup d'efforts sont maintenant consacrés pour trouver des inhibiteurs de son activité transcriptionnelle. L'activité transcriptionnelle de c-Myc passe par son hétérodimérisation avec Max, un autre facteur de transcription de type bHLH-LZ, et sa liaison à des régions promotrices d'ADN. Plusieurs groupes de recherche ont tenté de développer des petites molécules capables de lier c-Myc spécifiquement afin d'empêcher cette étape préalable. De par la nature intrinsèquement désordonnée de c-Myc, l'approche classique n'a pu permettre, jusqu'à présent, le développement d'inhibiteurs spécifiques avec une utilité clinique. Dans le cadre d'une précédente étude effectuée dans le laboratoire, nous avons montré qu'une région de Miz-1, un facteur de transcription de type BTB/POZ, pouvait lier spécifiquement c-Myc. Cette région de Miz-1, appelée Mid2, forme un complexe avec c-Myc et l'empêche d'hétérodimériser avec Max. Ainsi, le complexe c-Myc/Mid2 offre une plateforme alternative pour développer une nouvelle génération d'inhibiteurs contre c-Myc. Ce mémoire rapporte une étude structurale et biophysique du complexe c-Myc/Mid2 ainsi que sa capacité en tant qu'inhibiteur prototypique. Le chapitre 1 présentera un état des connaissances sur c-Myc et Miz-1, exposera les inhibiteurs de c-Myc déjà existants ainsi que les particularités attribuées aux protéines intrinsèquement désordonnées. Le chapitre 2 explique les méthodes utilisées afin de mener à bien le projet. Le chapitre 3 expose l'étude des caractéristiques biochimiques du Mid2 sous forme de petits peptides, ainsi que leur capacité à complexer c-Myc et de compétitionner avec Max afin d'empêcher la liaison de l'hétérodimère à l'ADN. L'application des déterminants structuraux élucidés dans ce mémoire de maîtrise permettront le développement d'une nouvelle génération d'inhibiteurs de l'oncoprotéine c-Myc, pouvant mener au final à de nouvelles thérapies pour vaincre le cancer.

Mots clés: c-Myc; Max; Miz-1; Addiction oncogénique; IDP; Dichroïsme Circulaire; Mid2; Cancer

SUMMARY

BIOPHYSICAL AND STRUCTURAL CHARACTERISATION OF THE C-MYC/MID2 COMPLEX – ELUCIDATION OF A MOLECULAR PLATFORM FOR NEW C-MYC INHIBITORS DISCOVERY

By
Jean-Michel Moreau
Programme de Biochimie

Master's thesis presented to the Faculté de Médecine et des Sciences de la Santé
For the diploma of Maître ès Sciences (M.Sc.) in Biochemistry
Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

The c-Myc oncoprotein, a bHLH-LZ type transcription factor, plays a central role in cellular growth and proliferation by activating and repressing a plethora of genes implicated in ribogenesis, metabolism, apoptosis and cell cycle. The importance of this role is highlighted when c-Myc expression levels are upscaled/deregulated by its translocation or the activation of pilot oncogenes. The protein then becomes oncogenic and contributes toward the tumorigenesis of a wide variety of cancers. Tumor cells overexpressing c-Myc are then addicted to its enhanced transcriptional activity. This addiction is the Achilles' heel of cancer cells. In fact, inhibition with a negative dominant protein (Omomyc) in a variety of cancer cell lines and mice models leads towards tumor regression. This work led to c-Myc being recognised as an important therapeutic target and considerable efforts are now being made to find inhibitors for its transcriptional activity. c-Myc's transcriptional activity is done in heterodimer with Max, another bHLH-LZ transcription factor, and their binding to DNA promoter regions. Many research groups attempted to develop small molecule inhibitors capable of specifically binding c-Myc to prevent the heterodimerization step. By the intrinsically disordered nature of c-Myc, rational drug design couldn't, to this day, develop specific inhibitors with a clinical utility. In previous experiments done in our laboratory, we showed that a region of Miz-1, a BTB/POZ transcription factor, could specifically bind c-Myc. This Miz-1 region, Mid2, forms a complex with c-Myc and prevents its heterodimerization with Max. Thusly, the c-Myc/Mid2 complex offers an alternative platform for the development of a prototypic c-Myc inhibitor. This Master's thesis is a structural and biophysical study of the c-Myc/Mid2 complex and its potency as a prototypic inhibitor. The first chapter will feature knowledge about c-Myc and Miz-1, existing direct c-Myc inhibitors and the particularities about intrinsically disordered proteins. Chapter two explains the methods used to carry out the project. Chapter three exposes a thermodynamical and biophysical study of Mid2 born peptides, their characterization in complex or not with c-Myc and their competition with Max to form the complex in presence of DNA. The application of elucidated structural determinants in this Master's thesis will permit the development of a new generation of direct inhibitors for the c-Myc oncoprotein, leading towards novel therapies to fight cancer.

Keywords: c-Myc; Max; Miz-1; Oncogene addiction; IDP; Circular Dichroism; Mid2; Cancer

Table des matières

Résumé	ii
Summary	iii
Table des matières	iv
Liste des figures	vi
Liste des abréviations.....	vii
Remerciements	xi
1. Introduction	12
1.1 Cancer	12
1.1.1 Définitions et prévalence.....	12
1.1.2 Tumorigénèse	13
1.2 Le proto-oncogène Myc et son importance dans le cancer	15
1.2.1 L'oncogène myc	15
1.2.2 Addiction oncogénique.....	18
1.2.3 Biologie moléculaire du réseau Myc/Max.....	19
1.2.4 Biologie structurale et thermodynamique du réseau Myc/Max	21
1.2.5 Approche classique de design d'inhibiteurs de Myc basée sur les molécules de synthèse	22
1.3 Miz-1	25
1.3.1 Biologie moléculaire de Miz-1	25
1.3.2 Interaction Myc/Miz-1	27
1.3.3 Compétition Miz-1/Max pour Myc	28
1.4 Protéines Intrinsèquement Désordonnées	29
1.4.1 Définitions et problématiques.....	29
1.4.2 Méthodes utilisées pour l'étude des IDPs.....	32
1.4.3 La structure Poly-L-Proline II	33
1.4.4 Les IDPs et les transitions de phase liquide-liquide	35
1.5 Objectifs de l'étude	36
2. Méthodes	37
2.1 Obtention des constructions protéiques	37
2.1.1 Clonage	37
2.1.2 Surexpression de protéines	38
2.1.3 Purification des protéines	40
2.2 Dichroïsme circulaire.....	41
2.2.1 Principes inhérents à la méthode	41
2.2.2 Utilisation du dichroïsme circulaire pour l'étude des protéines	41
2.2.3 Protocole d'utilisation du dichroïsme circulaire	43
2.3 Electrophoretic Mobility Shift Assays (Essais de retardement sur gel)	44
2.3.1 Principes inhérents à la méthode	44
2.3.2 Protocole d'utilisation des EMSAs.....	45
2.4 Résonance Magnétique Nucléaire à l'état liquide	45

2.4.1 Principes inhérents à la méthode	46
2.4.2 Spectres 1D ^1H -NMR	47
2.4.3 Spectres 2D ^1H - ^{15}N -HSQC	48
2.4.4 Titrations par ^1H - ^{15}N -HSQC	50
2.4.5 Protocole expérimental des titrations par ^1H - ^{15}N -HSQC	51
3. Résultats	53
3.1 Séquestration de c-Myc par les peptides Mid2s	53
3.1.1 Les peptides Mid2s- α , β et γ adoptent une conformation polyproline-II	53
3.1.2 Mid2s- β forme un complexe spécifique avec c-Myc plus stable que l'hétérodimère c-Myc/Max	56
3.1.3 Mid2s- β perturbe le complexe c-Myc/Max : ADN pour favoriser la formation du complexe Max/Max : ADN	59
3.1.4 Mid2s- β séquestre c-Myc pour favoriser l'homodimère de Max	65
3.1.5 Mid2s- β compétitionne avec Max pour former un complexe exclusif avec c-Myc	68
4. Discussion	73
4.1 Contexte et portée de l'étude	73
4.2 Discussion sur la séquestration de c-Myc par les peptides Mid2s	74
5. Conclusion	75
5.1 Perspectives	76
Liste des publications	76

Liste des figures

Figure 1. Comportement d'une cellule tumorale	13
Figure 2. Caractéristiques tumorales affichées par les déterminants de la cellule.....	14
Figure 3. Cartographie des domaines fonctionnels de c-Myc.....	15
Figure 4. Voies métaboliques et points de contrôles impliquant c-Myc.....	17
Figure 5. Inhibiteurs de c-Myc obtenus par criblage de petites molécules de synthèse	25
Figure 6. Cartographie des domaines fonctionnels de Miz-1	26
Figure 7. Énergie libre conformationnelle inhérente aux protéines globulaires et intrinsèquement désordonnées.....	30
Figure 8. Modes de repliement des protéines intrinsèquement désordonnées.....	32
Figure 9. Comparaison des structures secondaires classiques avec PPII.....	34
Figure 10. Fréquences de résonance de noyaux importants pour la spectroscopie RMN	46
Figure 11. ^1H -NMR de l'ubiquitine UbKEKS.....	47
Figure 12. ^1H - ^{15}N -HSQC de l'ubiquitine UbKEKS assignée.....	48
Figure 13. ^1H - ^{15}N -HSQC du bHLH-LZ de c-Myc assignée.....	49
Figure 14. Effets des types d'échanges sur les pics en ^1H - ^{15}N -HSQC	51
Figure 15. Caractérisation des peptides Mid2s- $\alpha/\beta/\gamma$ par dichroïsme circulaire	55
Figure 16. Caractérisation des complexes c-Myc/Mid2s/Max par dichroïsme circulaire	58
Figure 17. Titration du complexe c-Myc*/ADN avec Mid2s- β et modèle associé	63
Figure 18. Caractérisation des complexes c-Myc/Mid2s/Max par EMSA	64
Figure 19. Titration par ^1H - ^{15}N -HSQC du complexe c-Myc*/ ^{15}N -Max* avec Mid2s- β	67
Figure 20. Titration par ^1H - ^{15}N -HSQC du complexe ^{15}N c-Myc*/ ^{15}N -Max* avec Mid2s- β	71
Figure 21. Modèle général de séquestration du bHLH-LZ de c-Myc par Mid2s- β	72

Liste des abréviations

ADN	Acide désoxyribonucléique
AKT	Protein kinase B
ARN	Acide Ribonucléique
ARNm	ARN messenger
APC	Adenomatous polyposis coli
ARF	Alternate reading frame
ASLV	Avian Sarcoma Leukosis Virus
b-HLH-LZ	basic region helix-loop-helix leucine zipper
BTB	BR-C, ttk, bab
c-Myc	cellular Myc
c-Myc*	Le bHLH-LZ de c-Myc
Bcl-6	B-cell lymphoma 6 protein
BRD4	Bromodomain-containing protein 4
CD	Circular dichroism
ChIP-Seq	Chromatin immunoprecipitation-DNA Sequencing
DBD	DNA Binding Domain
DNase-I	Deoxyribonuclease-I
E-box	Enhancer box
EMSA	Electrophoretic Mobility Shift Assay
ERK	Extracellular Signal-Regulated Kinase
EGFR	Epidermal Growth Factor Receptor
FBW7	F-box and WD repeat domain-containing 7
FPLC	Fast Protein Liquid Chromatography
Fz	Frizzled
FOXO3a	Forkhead box O3a
FRET	Förster Resonance Energy Transfer
Gfi-1	Growth Factor Independent 1
GSK3 β	Glycogen Synthase Kinase 3 beta
HSQC	Heteronuclear Single Quantum Coherence
IDP	Intrinsically Disordered Protein
IPTG	Isopropyl β -D-1-thiogalactopyranoside
K-Ras	Kirsten-Ras
k_{on}	Constante de réaction
k_{off}	Constante de réaction inverse
K_d	Constante de dissociation
LB	Luria Broth
L-Myc	Leucocyte Myc
LLPS	Liquid-Liquid Phase Separation
Max	Myc-Associated factor X
MAPK	Mitogen-Activated Protein Kinase
MB	Myc Box
Miz-1	Myc-Interacting Zinc finger protein 1
MID	Myc Interacting Domain
mTORC1	mammalian target of rapamycin complex 1

Myc	avian Myelocytomatosis viral oncogene homolog
N-Myc	Neuroblastoma Myc
PDB	Protein DataBank
PEST	Proline-Acide Glutamique-Sérine-Thréonine
PI3K	Phosphoinositide 3-kinase
PIN1	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1
POZ	Pox virus and zinc finger
PP2A	Protéine Phosphatase 2A
PPII	Poly-L-Proline II
P-TEFb	Positive Transcription Elongation Factor
PTM	Post-Traductional Modification
SAXS	Small-angle X-ray scattering
SDS-PAGE	SodiumDodecylSulfate-PolyAcrylamide Gel Electrophoresis
S-Myc	Short Myc
SMAD	Small Worm Phenotype/Mothers Against Decapentaplegic
smFRET	small-molecule FRET
Ras	Rat Sarcoma
RMN	Résonance Magnétique Nucléaire
TAD	TransActivation Domain
Tet-off	Tetracycline-controlled transcriptional activation
TCF	T-Cell Factor
TCR	T-Cell Receptor
TGF β	Transforming Growth Factor Beta
TFIIH	Transcription Factor II Human
TopBP1	DNA topoisomerase 2-binding protein 1
TRRAP	Transformation/transcription domain-associated protein
TSS	Transcriptional Start Sites
VH	variable heavy
VL	variable light
V-Myc	viral Myc
Wnt	Wingless-Related Integrated Site
ZBTB17	Zinc finger and BTB domain-containing protein 17
ZF	Zinc Finger

*J'aimerais dédier ce mémoire à mes nièces Gaëlle et Mahée
Que vos rêves vous portent loin*

Scientific discovery and scientific knowledge have been achieved only by those who have gone in pursuit of it without any practical purpose whatsoever in view.

Max Planck

Remerciements

J'aimerais d'abord remercier tous les membres du laboratoire qui m'ont supporté durant mon parcours : Martin, Olivier, Vincent, Patrick, Danny, Laurent et Loïka. Les discussions scientifiques et surtout non-scientifiques qui ont ponctué mon passage au labo auront rendu l'expérience bien agréable.

La crew du Bac Marie-Anne, Gabriel, Dominic et Michael pour toute les activités improvisées mais toujours bien agréables.

Les boys de la plateforme Patrice, Étienne et Jean-Marie pour toute la bouille, etc. ainsi que toutes les belles rencontres que j'ai côtoyé durant mon passage à Sherbrooke.

Bien évidemment mes parents Lorraine et Michel, sans qui tout ce parcours n'aurait été possible. Un grand merci!

Finalement, un merci tout spécial à Pierre, pour avoir su rendre ce parcours bien plus qu'un apprentissage de la thermodynamique et de la RMN appliquée aux protéines, mais aussi une expérience humaine. Tu m'as également appris à naviguer et à voir le monde de mes yeux de scientifique, et ceci est un legs inestimable.

1. Introduction

Ce mémoire présente les travaux que j'ai effectué lors de ma maîtrise au laboratoire du professeur Pierre Lavigne.

1.1 Cancer

1.1.1 Définitions et prévalence

Le cancer est le nom donné à une collection de maladies apparentées. Dans tous les types de cancers, une ou des cellules se divisent sans arrêter, menant parfois à l'invasion de ces cellules dans le tissu atteint (NIH, 2019). Ces cellules en division incontrôlée forment des tumeurs : elles peuvent être solides ou liquides dépendamment du type de cancer. Lorsque les tumeurs atteignent une certaine taille, les cellules cancéreuses peuvent se détacher de la tumeur et envahir les tissus environnants via le sang ou le système lymphatique, causant l'apparition de métastases.

Le cancer est une maladie génétique, c'est-à-dire que l'apparition de la maladie est dû à un changement dans les gènes qui contrôlent la prolifération et la division cellulaire. Les mutations génétiques causant le cancer chez un individu peuvent être obtenues de manière héréditaire, ou bien apparaître spontanément durant la vie de l'individu. Ces mutations spontanées sont le résultat d'erreurs de réplication lors de la division cellulaire ou de dommages directs à l'ADN causés par des stress environnementaux qui seront mal réparés par la cellule, ou par le vieillissement. Le cancer a une incidence de plus en plus élevée avec l'âge; cela est causé par l'accumulation des risques de cancers au long de la vie en conjonction avec une perte d'efficacité des mécanismes de réparation dû au vieillissement.

Les stress environnementaux causant des dommages à l'ADN peuvent être réduits à trois grands types : les cancérogènes physiques, chimiques et biologiques (WHO, 2019). Les cancérogènes physiques les plus communs sont le rayonnement ultraviolet et les radiations ionisantes. Les cancérogènes chimiques comportent l'amiante, les composantes de la fumée du

tabac, l'arsenic, etc. Finalement, les cancérogènes biologiques englobent les infections dues à certains virus (e.g. Virus du Papillome Humain s'intégrant dans le génome) bactéries ou parasites.

Cinq types principaux de cancers ont été responsables de 8,8 millions de décès mondialement en 2015. Ces principaux types comportent le cancer du poumon (1,69 millions de décès), le cancer du foie (788 000 décès), le cancer colorectal (774 000 décès), le cancer de l'estomac (754 000 décès) et finalement le cancer du sein (571 000 décès). Le cancer est devenu en 2012 la première cause de décès au Canada, avec une proportion de 30,2% des cas (Canadian Cancer Society, 2012), venant souligner d'avantage l'importance de l'approfondissement des connaissances des mécanismes impliqués dans le cancer et comment les cibler efficacement

1.1.2 Tumorigénèse

Le terme néoplasme est utilisé pour signifier toute nouvelle croissance tissulaire anormale – leur nature peut être bénigne ou maligne. Le terme cancer est associé aux tumeurs malignes. Les cellules cancéreuses partagent certaines caractéristiques clés (Fig.1). Elles exhibent aussi une prolifération rapide en conjonction avec une réduction du contrôle de croissance, une perte d'inhibition de contact *in vitro*, un envahissement local des tissus et une dissémination vers d'autres parties du corps (métastases). Le pouvoir métastatique des cellules cancéreuses est le plus

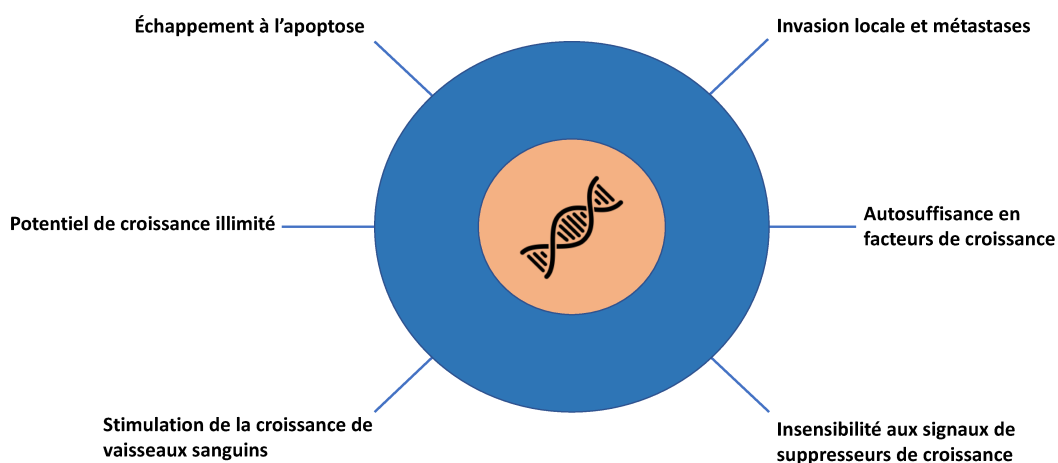


Fig.1 Comportement d'une cellule tumorale

Caractéristiques affichées par une cellule lorsqu'elle devient cancéreuse.

Adapté de Hanahan D, Weinberg RA. 2000

souvent responsable de la mort des sujets atteints de cancer. Les tumeurs bénignes quant à elles exhibent une diminution du contrôle de leur croissance, sans toutefois envahir les tissus locaux ou disséminer dans d'autres parties du corps (Hanahan D, Weinberg RA. 2000). L'initiation de la

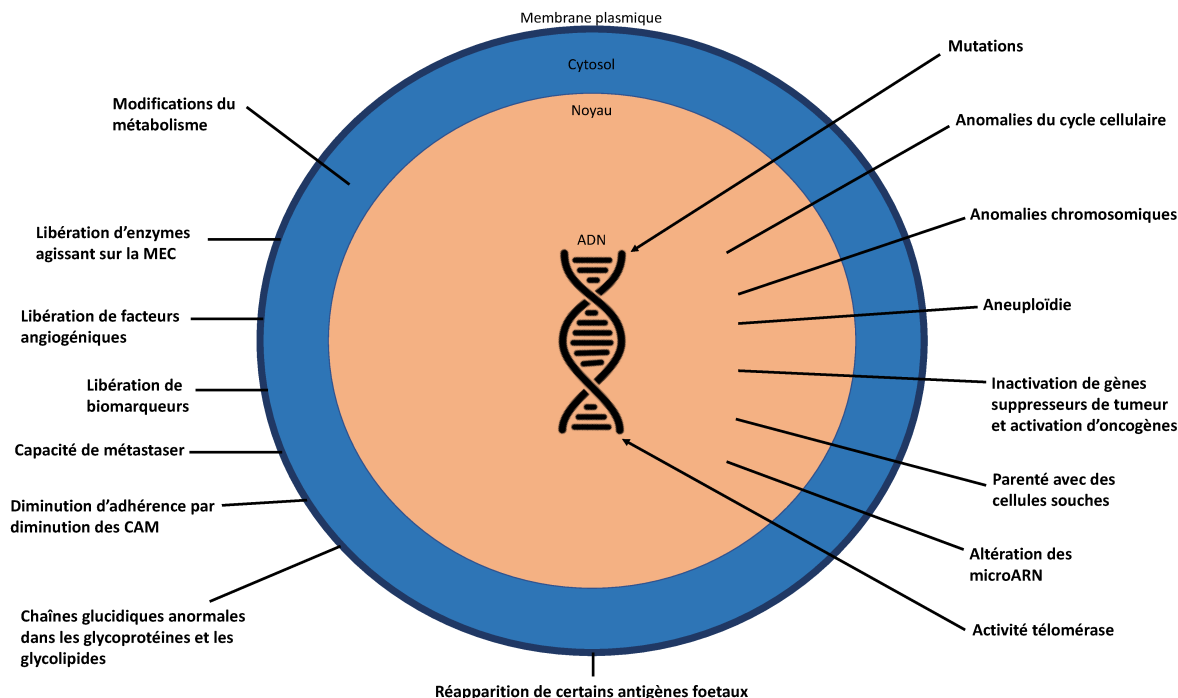


Fig.2 Caractéristiques tumorales affichées par les déterminants de la cellule
Modifications et mécanismes enclenchés par la cellule caractéristique de son état tumoral.
Adapté de Bender DA, 2013.

tumorigénèse correspond à une altération génétique non létale qui vient modifier le métabolisme général de la cellule (Fig. 2). Quatre classes de gènes peuvent causer l'apparition de tumeurs après altération génétique : les proto-oncogènes, les suppresseurs de tumeur, les gènes impliqués dans la réparation de l'ADN et les gènes pro-apoptotiques (Bender, DA. 2013). Les plus grandes questions en cancérologie restent encore l'élucidation des mécanismes biochimiques et génétiques qui mènent à la croissance incontrôlée des cellules ainsi que leur capacité d'invasion des tissus – c'est sur cette problématique générale que ce mémoire de maîtrise portera, plus précisément sur le proto-oncogène *c-myc*.

1.2 Le proto-oncogène Myc et son importance dans le cancer

1.2.1 L'oncogène myc

Le proto-oncogène a été mentionné pour la première fois dans la littérature dans une étude sur MC29, un virus causant la leucémie ainsi que diverses tumeurs comme des endothéliomes ou des carcinomes chez la volaille. Cette étude faisait mention du gène codant pour v-Myc (viral Myc). Le gène a été nommé ainsi selon l'*avian myelocytomatosis viral oncogene homolog* (Dalla-Favera R, et al. 1982). Afin d'approfondir les connaissances sur l'effet de Myc, des mutants de MC29 ont été générés, et étaient capables de transformer les fibroblastes pour induire des tumeurs *in vivo*. Toutefois, ces mutants montraient une capacité fortement diminuée à transformer les cellules hématopoïétiques pour induire la formation de tumeurs. Une étude biochimique de ces mutants révèle alors que les peptides spécifiques à l'oncogène myc n'étaient plus exprimés, ce qui explique la capacité diminuée de formation de tumeurs (Ramsay, G. et al. 1982, Enrietto, P. J. et al. 1982). Il fut alors déterminé, suite à l'analyse de l'ARN viral, que ces mutants présentaient des délétions importantes dans l'oncogène myc, venant ainsi montrer que myc était responsable des propriétés oncogéniques du virus MC29 *in vivo* (Bister, K. 1982). Plusieurs homologues de v-Myc ont été découverts par la suite, tels que N-Myc (neuroblastoma), L-Myc (leucocyte) (Kohl et al. 1983) et S-Myc (short) (Sugiyama et al. 1989). L'orthologue cellulaire de v-Myc, c-Myc (ci-après

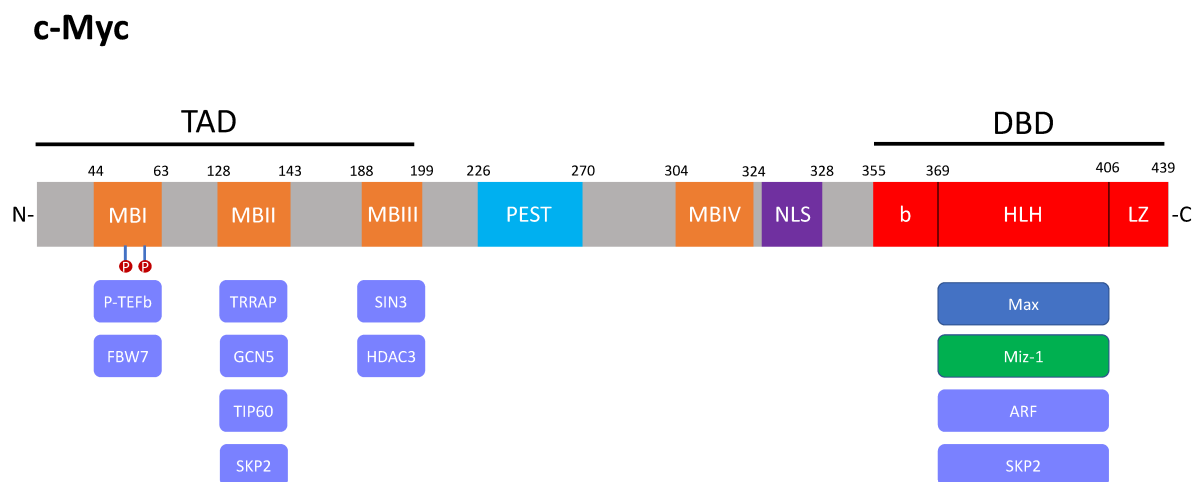


Fig.3 Cartographie des domaines fonctionnels de c-Myc

Domaines fonctionnels positionnés sur la structure primaire de c-Myc, de N à C terminal.

Adapté de Carabet LA, 2018.

Myc), est situé sur le locus 8q24.21 (6 kilobases) et possède trois exons, dont le premier est non codant. La transcription de Myc est régulée par quatre promoteurs distincts (P_0 , P_1 , P_2 et P_3). Le gène contient deux codons d'initiation (le second codant pour S-Myc) principaux (ATG et CTG), deux sites de polyadénylation et plusieurs sites hypersensibles à la DNase-I (Meyer N, et al. 2008). La protéine provenant de ce gène (439 acides aminés) est constituée d'un site de transactivation en N-terminal (TAD – *TransActivation Domain*), et d'un site de liaison à l'ADN de type bHLH-LZ (*basic region Helix-Loop-Helix – Leucine Zipper*) en C-terminal liant préférentiellement les séquences E-box (CACGTG). Le tout contient quatre « Myc Boxes » (MB I à IV), des sites de phosphorylation et de liaison protéine-protéine, ainsi qu'une région riche en proline, acide glutamique, sérine et thréonine nécessaire pour une dégradation rapide (PEST) (*Fig. 3*) (Meyer N, et al. 2008).

Myc a d'abord été caractérisée pour son implication dans l'apparition de lymphomes B chez le poulet, après intégration du rétrovirus ASLV (*Avian Sarcoma Leukosis Virus*) près du locus de Myc stimulant son expression (Hayward et al. 1981). Dans le cas des lymphomes de Burkitt chez l'humain, la translocation t(8;14) déplace le gène de Myc (normalement situé à 8q24) au locus de la chaîne lourde (VH) des Immunoglobulines sur le chromosome 14. L'expression du gène de Myc se retrouve alors contrôlé par un promoteur fortement actif résultant alors en une stimulation de l'expression du gène (Taub et al. 1982, Adams et al. 1983). D'autres translocations sont rapportées pour d'autres types de tumeurs – t(8;2) et t(8;22), liant le gène de Myc au locus de la chaîne légère des Immunoglobulines (VL) (Bernard, O. et al. 1983).

L'oncoprotéine Myc se retrouve surexprimée ou s'accumule par faute de dégradation dans une grande majorité de cancers même si son gène n'est ni transloqué, ni muté. Sa surexpression et son accumulation (section 1.2.3) dépendent de l'activation d'oncogènes moteurs 'driver oncogene' en aval ou de l'inactivation de suppresseurs de tumeurs. Comme nous le verrons plus bas, cette surexpression ou accumulation rend les cellules tumorales dépendantes aux activités transcriptionnelles de c-Myc. En temps normal, une surexpression de Myc dans des cellules activera des points de contrôle empêchant la prolifération tels que Arf ou p53 (*Fig.4*) (Dang, C. 2012). Le potentiel tumorigénique du facteur de transcription se révèle lorsque ces points de contrôle ne sont plus actifs via une mutation enlevant leur fonctionnalité, ce qui est le cas dans

bien des lymphomes (Eischen et al. 1999). Le potentiel tumorigénique de Myc peut aussi être engagé via une mutation constitutivement active dans une voie de signalisation telle que K-ras ou Wnt (D'Cruz et al. 2001). Autre exemple, des défauts dans la voie Wnt-APC (mutation dans le suppresseur de tumeur APC) dans le carcinome colorectal humain résultent en une activation transcriptionnelle élevée de Myc via TCF. Dans les lymphomes, Myc se retrouve en aval de la voie de signalisation Notch dérégulée (Dang, C. 2012). Il est à noter que la surexpression de Myc doit dans certains cas être associée à de nouvelles mutations dans les voies de signalisation pour causer l'apparition de tumeurs.

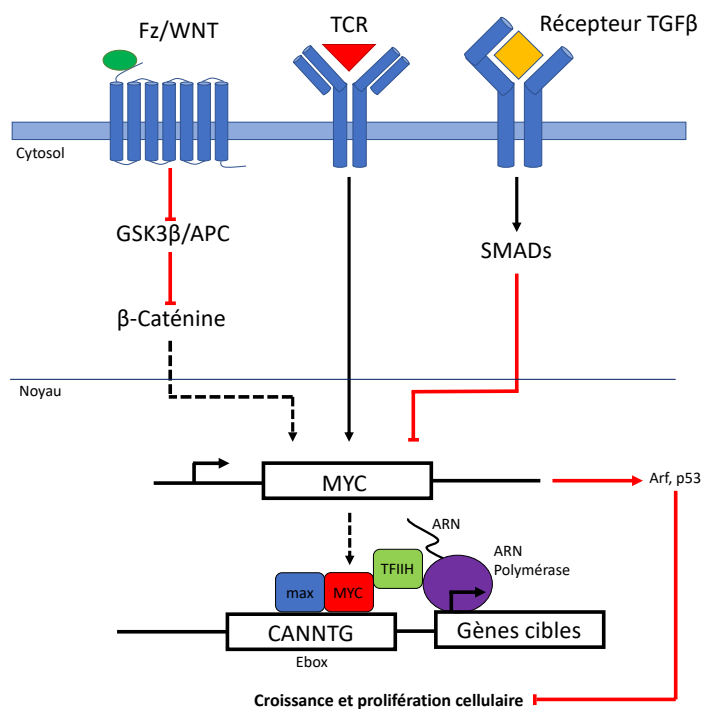


Fig.4 Voies métaboliques et points de contrôles impliquant c-Myc

Voies métaboliques importantes dans le contrôle de la transcription de c-Myc (Fz/Wnt, TCR, TGFβ)

Adapté de Dang C, 2012.

Il est convenu, après 40 ans d'études sur Myc, que cet oncogène est impliqué d'une manière ou d'une autre dans au moins 75% des cancers humains (Carabet, L. et al. 2018). Cela inclut principalement, dans les cancers solides, le cancer de la prostate, du sein, du colon et du cerveau. Pour les cancers non-solides, on retrouve la leucémie myéloïde, les lymphomes de Burkitt, les carcinomes du poumon ainsi que les neuroblastomes (Kalkat, M. et al. 2017). L'implication quasi-universelle de Myc dans le cancer rend l'oncogène une cible de choix pour contrer cette pathologie.

1.2.2 Addiction oncogénique

L'addiction de la cellule à un oncogène ne provient pas seulement que de l'accumulation de mutations activant la transcription d'oncogènes et l'inactivation de gènes suppresseurs de tumeurs, elle provient du fait que ces gènes sont impliqués dans des voies de signalisation complexes et centrales aux cellules. Les cellules cancéreuses se retrouvent alors avec une signalisation propre à leur état transformé, qui est unique en son genre et qui ne s'apparente en rien à une cellule saine (Weinstein, IB. 2002). Ces cellules cancéreuses se retrouvent alors dépendantes aux oncogènes maintenant leurs voies de signalisations uniques, sans quoi elles entreraient en sénescence ou en apoptose. En simples termes, l'addiction de la cellule à un oncogène se caractérise le plus souvent par la maintenance d'un phénotype tumorigénique malin causé seulement par un ou quelques gènes. (Weinstein, IB. 2008). Ce principe a été démontré avec c-Myc à l'aide de modèles murins ayant des cellules à expression *tet-off* pour l'oncogène Myc. Les souris étaient injectées avec ces cellules tumorales (dans ce cas-ci, des ostéoblastes causant le cancer des os) Myc dépendantes. Un groupe de souris n'a pas été traité à la doxycycline afin de garder l'expression de Myc active en fonction du système *tet-off*. Deux autres groupes ont été traités à la doxycycline pendant toute la durée de l'étude (4 mois) et pendant seulement 10 jours, respectivement. Le groupe de souris ayant été traités durant toute la durée de l'étude n'as pas développé de tumeurs, alors que 85% du groupe ayant été partiellement traité sont mortes après 3 mois des suites de tumeurs, et 100% du groupe de souris n'ayant pas été traité à la doxycycline sont mortes après 1 mois pour le même motif. Cette étude montre que les tumeurs progressant sous le contrôle de niveau élevée de c-Myc s'effondrent lors de son retrait. Les tissus murins extraits montrent une différenciation des ostéoblastes tumoraux vers des ostéocytes, ainsi qu'une régression notable des tumeurs témoignant de l'apoptose des cellules tumorales. (Jain, M. et al. 2002). L'addiction oncogénique de la cellule à Myc est donc l'un des talons d'Achille de la vaste majorité des cancers. C'est pour cette raison que cet oncogène est une cible de choix.

1.2.3 Biologie moléculaire du réseau Myc/Max

Myc est un facteur de transcription de type b-HLH-LZ activateur de croissance et de prolifération cellulaire. Les niveaux physiologiques (dans une cellule saine) de ce facteur de transcription sont augmentés en fonction de stimuli mitogéniques ou diminués en retour de stimuli inhibiteurs de croissance (par exemple Arf et p53). Myc, de par sa nature de protéine intrinsèquement désordonnée (informations supplémentaires en 1.4), doit s'associer à des partenaires d'interaction afin d'effectuer son rôle de régulation dans la cellule. Ainsi, pour lier l'ADN, Myc interagit avec son partenaire dit 'obligatoire' Max, une autre protéine de type bHLH-LZ. Ce dimère régulateur de croissance cellulaire lie préférentiellement les séquences E-box (CACGTG) sur l'ADN. Fait particulier, ce dimère peut réprimer la transcription de gènes importants pour l'arrêt du cycle cellulaire, p15^{INK4B} et p21^{CIP1}, des facteurs de transcriptions tels que Miz-1, ainsi que des inhibiteurs de kinase cycline-dépendants (Seoane J, et al. 2002). Dans des cas mentionnés en 1.2.1 où l'on retrouve Myc surexprimés dans la cellule, cette répression de gènes cyostatiques peut être très problématique – puisque la croissance cellulaire et la prolifération sont trop fortement encouragés. Dans une cellule saine cependant, le TGFβ (*Transforming Growth Factor β*), une cytokine antimitogénique, permet via la voie de signalisation des SMADs une répression de la transcription de Myc (Fig. 4). Étant donné que l'ARNm de Myc ainsi que la protéine associée n'ont qu'une survie cellulaire d'environ 30 minutes avant dégradation, la simple inhibition de transcription de Myc via TGFβ suffit à contrer les effets pro-prolifératifs associés à ce facteur de transcription (Seoane J, et al. 2001).

La survie cellulaire de 30 minutes de Myc est régulée principalement via des signaux mitotiques provoquant des phosphorylations dans la MBI (Fig.3). Un signal mitotique provenant d'EGFR (voie des MAPK) actif, va permettre via une activation de Ras et Erk la phosphorylation de la sérine 62 sur Myc en même temps d'inhiber la voie PI3K/AKT. Lors du retrait du signal mitotique, par exemple lorsque les niveaux de Myc endogènes sont trop élevés, l'inhibition de PI3K est levée et permet via GSK3 la phosphorylation de la thréonine 58. Cette double phosphorylation dans la MBI est un motif de reconnaissance pour PIN1, qui vient changer la conformation de Myc pour permettre à la protéine phosphatase 2A (PP2A) de déphosphoryler la S62. Une T58 phosphorylée seule dans la MBI est un motif de reconnaissance pour FBW7 (Fig.

3) qui va permettre la polyubiquitination sur la Lys 48 et une dégradation de Myc. (Yeh E, et al. 2004) (Farrell AS, Sears RC, et al. 2014).

Cependant, certaines dérégulations associées à la tumorigénèse peuvent venir interférer dans la survie cellulaire courte de Myc, et lui permettre de persister. Prenons comme exemple une mutation dans la voie des MAPK, EGFRviii, qui rend le récepteur constitutivement actif (Tompa M, et al. 2019). Ce signal mène à la phosphorylation de la S62. Seulement, une activation constante de EGFR amène une inactivation permanente de la voie PI3K/AKT, empêchant alors GSK3 de phosphoryler la T58. Myc ne peut alors subir les modifications nécessaires à sa dégradation – menant donc à sa persistance dans la cellule (Yeh E, et al. 2004). Une telle accumulation peut mener à une présence de Myc à des niveaux 10 000 fois supérieurs aux niveaux observés de manière endogène. Une présence si grande de Myc mène à une expression fortement augmentée de ses gènes cibles (pour la plupart pro-prolifératifs) ainsi qu'à une invasion de gènes non-spécifiques – des éléments clés pour la tumorigénèse. Il est à noter que Myc, dans ce cas-ci, n'a pas subi de mutations – il se retrouve dans le rôle d'oncogène moteur suite à une/des mutations dans des voies de signalisation. Le cas présenté ici utilise la voie des MAPK, mais le même scénario pourrait se produire avec d'autres voies de signalisation.

De manière intéressante, les gènes ciblés ne se chevauchent pas dans les différents types de cancer Myc-dépendants (Chandriani S, et al. 2009). Pour cette raison, il est difficile d'assigner une certaine signature oncogénique propre à Myc. Les quelques gènes communs à la signature oncogénique de Myc sont reconnus comme étant impliqués dans les mécanismes de métabolisme basal et de prolifération de la cellule (Ji H, et al. 2011). Cependant, cette signature commune n'explique pas la majorité des effets oncogéniques observés dans les cancers Myc-dépendants. Il a été démontré que dans les cellules tumorales surexprimant Myc, le facteur de transcription lie les gènes les plus actifs et augmente leur expression – et non pas un panel de gènes défini. En d'autres termes, la surexpression de Myc mène à une amplification du programme d'expression génique existant et cette amplification devient tumorigénique et addictive (Lin CY, et al. 2012). Bien qu'il n'existe aucune signature commune, il existe un consensus sur le fait que la surexpression de gènes métaboliques, pro-prolifératifs et anti-apoptotiques rend les cellules

tumorales dépendantes à c-Myc et leur procure un avantage prolifératif en comparaison avec les cellules saines.

Des expériences de ChIP-Seq viennent supporter les hypothèses mentionnées ci-haut sur l'invasion de Myc sur les promoteurs non-spécifiques ainsi que l'amplification du programme d'expression génique tumorale. Une invasion du génome par Myc sur la chromatine transcriptionnellement active a été rapportée par cette méthode, dans le cas de cellules cancéreuses (Rahl PB, et al. 2010) (Lin CY, et al. 2012). À mesure que la concentration cellulaire de Myc croît, de plus en plus de promoteurs et d'amplificateurs situés sur la chromatine transcriptionnellement active se retrouvent liés par Myc, sans être influencés par la présence ou non de séquences E-box. Qui plus est, l'endroit où est retrouvé Myc sur la chromatine est corrélé avec la localisation de la polymérase II, plutôt que les séquences E-box (Guo J, et al. 2014). Il a d'ailleurs été démontré que les dimères tels que Myc-Max lient les séquences d'ADN non-spécifiques avec seulement un à deux ordres de grandeur de moins en affinité que les séquences E-box (Sauvé S, et al. 2007; Guo J et al. 2014).

Lorsque Myc lie l'ADN sous forme d'hétérodimère avec Max sur un promoteur, un complexe d'activation de la transcription est recruté par le domaine TAD de Myc. Le facteur d'élongation, P-TEFb et la protéine à bromodomaine BRD4 (Fig.3) sont recrutés afin de former un complexe qui permettra de phosphoryler la queue de la polymérase II. Cette phosphorylation lève les pauses de cette dernière et active la transcription de gènes cibles, ou dans le cas de cellules cancéreuses, de gènes non spécifiques. Dans le cas où l'hétérodimère Myc-Max lierait un amplificateur, le domaine TAD engagerait le recrutement du complexe TRRAP, un complexe protéique permettant d'ouvrir la chromatine via une acétylation des histones, la rendant transcriptionnellement active (Bradner, J. et al. 2016).

1.2.4 Biologie structurale et thermodynamique du réseau Myc/Max

Comme mentionné précédemment, Myc et Max sont des protéines de la famille des bHLH-LZ. Les protéines appartenant à cette catégorie dimérisent via les domaines HLH et leucine-zipper pour former une hélice double sous forme de superhélice (*coiled coil*) (O'Shea EK, et al. 1991).

Les hélices doubles adoptant cette forme sont caractérisées par une répétition en heptade $(abcdefg)_n$ où se positionnent des résidus hydrophobes (a et d) et des résidus chargés (e et g). Si les résidus de deux hélices opposées ont des charges complémentaires en e et g , à une heptade d'intervalle, la formation de ponts salins interhélices ($i, i'+5$) est favorisée. Ces ponts salins viennent stabiliser l'interaction en hélice double du leucine zipper, et sont retrouvés dans les homodimères de Myc, de Max et les hétérodimères Myc-Max, entre autres (Lavigne P, et al. 1995). Malgré la stabilité offerte par la formation des *leucine zipper*, les protéines de la famille des bHLH-LZ nécessitent à la fois le domaine HLH et le LZ pour former un dimère stable et une reconnaissance moléculaire efficace (e.g. l'hétérodimère Myc-Max) (Amati B, et al. 1993).

Myc peut former des homodimères, mais seulement de manière faible. Deux résidus Glu et un résidu Arg positionnés en a des heptades favorisent la dimérisation alors que six chaînes latérales de résidus chargés se retrouvent à l'interface du dimère, défavorisant l'interaction (Lavigne P, et al. 1995). Myc se retrouve en solution sous un ensemble de conformations plutôt que sous forme d'homodimère, de par sa nature intrinsèquement désordonnée (voir section 1.4 pour les IDPs).

1.2.5 Approche classique de design d'inhibiteurs de Myc basée sur les molécules de synthèse

Comme mentionné dans les précédentes sections, Myc présente une cible thérapeutique de choix pour traiter le cancer (voir 1.2.1). Malgré le nombre extrêmement élevé d'études publiées sur environ 40 ans, un inhibiteur direct de l'oncogène Myc reste encore hors de portée – d'où le qualificatif de plus en plus associé à la protéine : *undruggable* (Horiuchi D, et al. 2014). Le problème persistant avec Myc vient de sa nature intrinsèquement désordonnée, qui rend le design rationnel d'inhibiteur basé sur la structure difficile. Cependant, de nombreux groupes de recherche persistent néanmoins à atteindre cet objectif.

Différentes stratégies ont été considérées afin d'inhiber les effets oncogéniques de Myc. Ces stratégies incluent sa transcription, la stabilité de son ARNm, la stabilité de la protéine et sa dégradation, et la liaison de Myc à son interactome (un inhibiteur direct) (Whitfield JR, et al. 2017). Des exemples plus spécifiques de ces inhibiteurs seraient des stabilisateurs de G-quadruplex

empêchant Myc d'initier la transcription, des oligonucléotides antisens induisant la dégradation de l'ARNm de Myc, etc. Les cibles régulatrices indirectes visées sont principalement GSK3, Ras/MAPK, PP2A et FBW7. Les inhibiteurs de voies métaboliques permettant la traduction de Myc visent principalement MAPK, mTORC1 et FOXO3a (Koh CM, et al. 2016).

À ce jour, aucun médicament sous forme de petite molécule empêchant l'interaction Myc-Max ou leur liaison à l'ADN n'a été approuvée pour un traitement chez l'Homme. Il est à noter que les composés prototypiques développés en vue du traitement sont ressortis d'essais de criblage sur une librairie limitée de molécules. Il serait plus efficace de faire du design rationnel d'inhibiteurs basé sur la structure. Mais comme déjà mentionné, la structure de Myc est évasive, ce qui vient limiter les possibles efforts à cet égard. Les travaux présentés dans ce mémoire s'attaquent d'ailleurs à ce problème (Section 1.4).

Deux catégories d'inhibiteurs prototypiques provenant de librairies de molécules existent. Il y a les inhibiteurs interférant avec l'hétérodimérisation Myc-Max, et ceux bloquant l'interaction de l'hétérodimère Myc-Max sur l'ADN. Le but des inhibiteurs directs de Myc est de réduire son abondance biodisponible et ainsi réduire la prolifération et la croissance cellulaire, dans l'optique de renverser l'addiction oncogénique et induire l'apoptose. (Weinstein IB. 2002) (Carabet LA, et al. 2018). Ce mémoire portera une attention plus particulière aux inhibiteurs empêchant l'hétérodimérisation Myc-Max, plutôt que ceux empêchant la liaison à l'ADN, puisque le projet faisait l'objet du présent manuscrit cible spécifiquement l'hétérodimérisation Myc-Max.

Le premier inhibiteur prototypique rapporté du type d'inhibition de l'hétérodimérisation Myc-Max, le IIA6B17 (Fig.5), a été découvert parmi une librairie de 7000 peptidomimétiques en effectuant des expériences FRET (Förster Resonance Energy Transfer). Ce composé permettrait de supprimer la croissance de fibroblastes tumoraux induits par Myc. Cependant, une inhibition de Jun, un facteur de transcription, a été détectée, ce qui suggérerait une mauvaise spécificité du composé. De plus, aucun signal d'apoptose n'a été rapporté (Berg T, et al. 2002). Une autre équipe de recherche a apporté quelques modifications structurales à cette librairie de peptidomimétiques, et a identifié deux composés optimisés pour l'inhibition de Myc : Mycmycin-1 et Mycmycin-2.

Ces molécules avaient l'avantage de ne pas inhiber Jun tout en empêchant l'effet oncogénique de Myc (Shi J, et al. 2009).

Le groupe de recherche de E.V. Prochownik, en utilisant la librairie de composés Chembridge DIVERSet comprenant 10 000 molécules et en effectuant des expériences de double-hybride, a détecté sept nouveaux inhibiteurs potentiels. Ces inhibiteurs bloquaient la formation de l'hétérodimère Myc-Max à l'interface du HLH-LZ et ne semblaient pas toxiques *in cellulo*. De ces tests initiaux, trois composés ont été sélectionnés de par leur spécificité pour Myc-Max pour une étude plus approfondie : 10058-F4, 10074-G5 et 10074-A4 (Fig.5). Quoiqu'ayant eu un succès initial dans des essais en cellule, ces composés présentaient une efficacité faible (e.g. $IC_{50} = 41,1 \mu M$ pour 10058-F4, $IC_{50} = 22,5 \mu M$ pour 10074-G5) sur la croissance de cellules cancéreuses Myc dépendantes HL-60 (Wang H, et al. 2013). Ces trois molécules ont une application clinique limitée puisqu'ils présentent une mauvaise sélectivité, un comportement pharmacocinétique peu convenable et sont dégradés rapidement, résultant alors en une faible concentration tumorale (Guo J, et al. 2009). Une refonte des ces molécules était alors nécessaire. En se basant sur 10058-F4, une recherche de similarité assistée par ordinateur a été effectuée sur plus de 500 000 composés de la Chembridge database, mais les tests *in vitro* / *in cellulo* ne montraient pas d'amélioration notable (Wang H, et al. 2007). La refonte de 10074-G5 a donné naissance à un analogue (JY-3-094) (Fig.5) via des études de SAR (*Structure-Activity Relationship*). Cet analogue montrait une meilleure capacité à inhiber l'hétérodimérisation Myc-Max *in vitro*, mais un problème de pénétration cellulaire persiste (Yap JL, et al. 2013). Finalement, le groupe de recherche a produit un congénère de 10074-G5, 3jc48-3 (Fig.5), dans une autre tentative d'améliorer leur composé. Les essais initiaux montrent une efficacité cinq fois plus élevée que 10074-G5, et une sélectivité deux fois plus élevée pour l'hétérodimère Myc-Max que pour l'homodimère Max-Max. Dans des cellules HL-60, 3jc48-3 induit l'arrêt de croissance cellulaire à la phase G0/G1 et possède une demi-vie cellulaire >17 heures. À ce jour, 3jc48-3 est le composé le plus efficace, actif en cellule et stable que le groupe de recherche de E.V. Prochownik a produit (Chauhan J, et al. 2014).

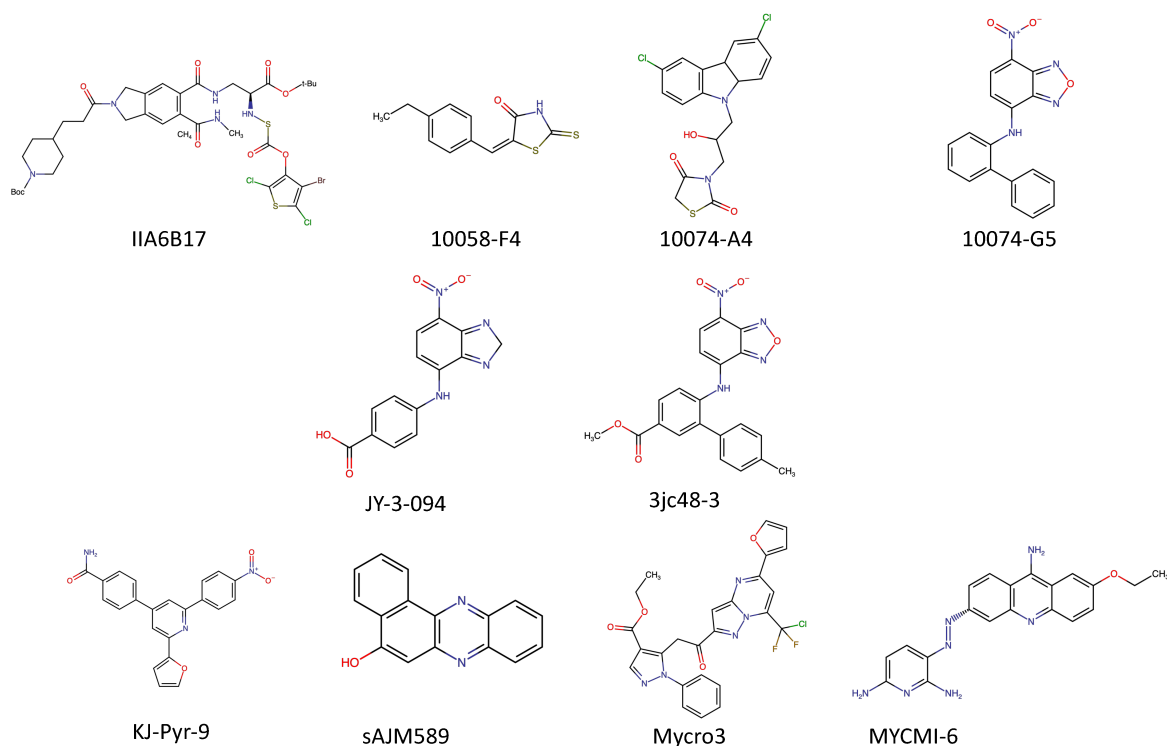


Fig.5 Inhibiteurs de c-Myc obtenus par criblage de petites molécules de synthèse
 Inhibiteurs peptidomimétiques de c-Myc obtenus par criblage à haut débit contre différentes librairies de molécules.
 Adapté de Carabet LA, 2018.

Plusieurs autres inhibiteurs prototypes de Myc dans cette catégorie existent, tels que Mycro3, KJ-Pyr-9, sAJM589 et MYCMI-6 (Fig.5), mais possèdent tous une biodisponibilité relativement faible ainsi que des efficacités et stabilités insuffisantes (Carabet LA, et al. 2018).

1.3 Miz-1

1.3.1 Biologie moléculaire de Miz-1

La protéine Miz-1 (*Myc-interacting Zinc finger protein-1*), également appelée ZBTB17, est un facteur de transcription de type BTB/POZ de 83 kDa pouvant former des complexes avec entre autres Myc, Gfi-1 et Bcl-6 pour effectuer son activité biologique. (Luo J, et al. 2016). Ce facteur de transcription est un activateur de gènes cytotatiques, spécifiquement des inhibiteurs de kinase cycline-dépendants tels que p15^{INK4B}, p21^{CIP1} et p57^{KIP2} (Bédard M, et al. 2016; Eilers M, Staller P, et al. 2001). Certaines séquences reconnues par ce facteur de transcription contrôlent des

gènes impliqués dans l'autophagie, liant l'inhibition du cycle cellulaire par Miz-1 à l'autophagie (Wolf E, et al. 2013). Miz-1 est exprimée de manière ubiquitaire durant le développement et peut fonctionner en tant qu'activateur de transcription et aussi comme répresseur de transcription dépendamment de son partenaire d'interaction (Wanzel M, et al. 2003). Miz-1 active p15^{INK4B} via un complexe avec Smad3/Smad4 par le recrutement de p300, le tout en réponse à un signal de TGFβ (Seoane J, et al. 2002). Cependant, il a été observé que Miz-1 possède une forte activité de répression de croissance cellulaire qui n'est pas associée à la voie TGFβ (Herold S, et al. 2002). Le domaine POZ de Miz-1 est reconnu comme étant un domaine régulant la majorité des interactions de la protéine et a d'abord été identifié comme le domaine expliquant la forte activité de répression de la croissance cellulaire. L'expression de p21^{CIP1} est fortement augmentée par Miz-1 lors de stimuli associés aux dommages à l'ADN. Cependant, le complexe Myc/Miz-1 régule négativement l'expression du gène cytotatique, lorsque les niveaux cellulaires de Myc sont plus élevés que ceux de Miz-1.

Il a été démontré qu'une stabilisation de Miz-1 à l'aide d'inhibiteurs d'ubiquitine ligases spécifiques (HUWE1) peut inhiber l'activation de gènes Myc dépendants et la progression du cycle cellulaire. En d'autres mots, cela suggère que des niveaux de Miz-1 plus élevés que ceux de Myc mènent à la répression des gènes pro-prolifératifs activés par Myc. De plus, des expérimentations en ChIP-Seq montrent une augmentation des niveaux de Miz-1 aux promoteurs liés par Myc près des *Transcriptional Start Sites* (TSS). Il est d'ailleurs montré que le ratio Myc/Miz-1 aux TSS dicte si un gène régulé par Myc est transcrit ou non. À un niveau plus élevé de Myc par rapport à

Miz-1

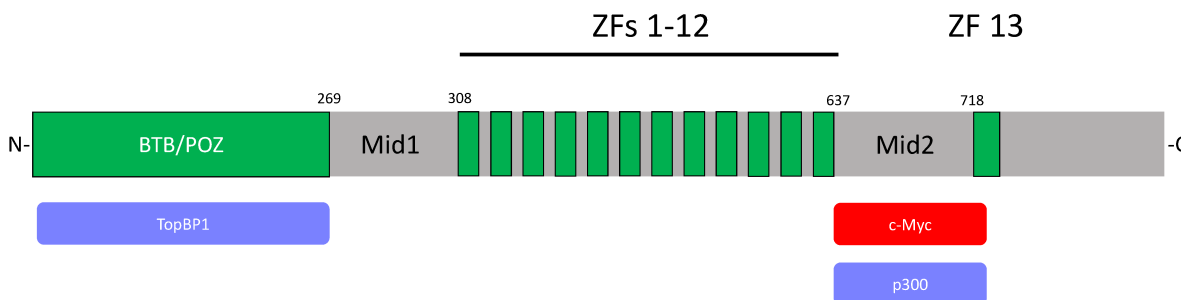


Fig.6 Cartographie des domaines fonctionnels de Miz-1

Domaines fonctionnels positionnés sur la structure primaire de Miz-1, de N à C terminal.

Adapté de Bédard M, 2016.

Miz-1, les gènes Myc-dépendants sont activés alors qu'à un niveau égal ces gènes sont réprimés (Peter S, et al. 2014).

À la suite de son domaine BTB/POZ en N-terminal, Miz-1 possède 13 doigts de zinc, dont 12 consécutifs. Les domaines permettant la liaison de Myc à Miz-1, les MIDs (*Myc Interacting Domain*), sont présents entre le BTB/POZ et les ZFs (MID1) et entre le ZF12 et le ZF13 (MID2) (Fig.6) (Bédard M, et al. 2016). Le domaine BTB/POZ a été caractérisé comme étant le principal domaine d'interaction de la protéine. À titre d'exemple, lors de dommages à l'ADN (e.g. exposition de la cellule à l'UV) l'interaction entre TopBP1 et Miz-1 via le domaine BTB/POZ est levée, libérant alors le domaine et permettant donc l'activation de p15^{INK4B} – amenant l'arrêt du cycle cellulaire (Herold S, et al. 2002). Miz-1 dimérise via son domaine BTB/POZ en structure cristalline (Stogios PJ, et al. 2019). La troncation de ce domaine rend la protéine non fonctionnelle, jusqu'à une certaine mesure, et empêche la liaison du facteur de transcription à la chromatine (Kosan C, et al. 2010). Il a aussi été déterminé que cette troncation permet à Miz-1 d'échapper à la régulation inhibitrice indirecte par Myc. Un mutant Miz-1 Δ POZ a été utilisé comme appât dans une expérience de double-hybride, et l'interaction Myc/Miz-1 était toujours présente, montrant que ce domaine n'est pas nécessaire pour l'interaction. (Peukert K, et al. 1997). Cependant, Miz-1 est de plus en plus étudiée pour ces autres domaines fonctionnels, tels que les doigts de zinc et leur rôle dans la topologie de l'ADN, ainsi que le rôle des régions Mid1 et Mid2 (Bédard M, et al. 2016).

1.3.2 Interaction Myc/Miz-1

La protéine Miz-1 a d'abord été identifiée comme une protéine interagissant avec Myc dans une expérimentation en double-hybride. Il a été déterminé que la délétion des a.a. 269-308 (Mid1) ou des a.a. 637-718 (Mid2) abolissait l'interaction entre Myc et Miz-1 (Fig.6). Une prédiction de structure secondaire prévoyait la présence d'hélice- α amphiphile dans la région Mid2 (Peukert K, et al. 1997). Étant donné que le HLH de Myc adopte une structure similaire (Ferré-D'Amaré AR, et al. 1993), il est convenu que l'interaction détectée est due à cette similarité de structure. Dans le but de défaire cette interaction, différentes constructions de Miz-1 ont été effectuées et testées contre Myc en essais double-hybride. En premier lieu, un mutant L685P,

I699P, A702P, V703G et V706P nommé « 4 Pro » a abolit l'interaction entre Miz-1 et Myc. En second lieu, un mutant I699P et S700G nommé « 1 Pro » a réduit l'affinité entre Miz-1 et Myc d'un ordre de 10. Ces expérimentations viennent alors circonscrire le site d'interaction de Myc sur Miz-1 – l'interaction entre les deux facteurs de transcription se passe principalement via la région Mid2 et le HLH de Myc, le tout, supposément, en hélice- α (Peukert K, et al. 1997).

Comme mentionné brièvement ci-haut, Miz-1 joue un rôle important dans l'arrêt du cycle cellulaire via l'activation de gènes cytotostatiques. Myc est un inhibiteur direct de cette activation via son interaction avec le domaine Mid2 sur Miz-1. Miz-1 nécessite la liaison d'un coactivateur, p300, sur la région Mid2 afin d'effectuer sa transactivation. Cependant, si Myc s'y trouve, la coactivation ne peut avoir lieu. Il est connu que la région 683-715 (correspondant à la partie C-terminale de Mid2) est nécessaire pour l'interaction Miz-1/p300, signifiant que p300 et Myc ne compétitionnent pas pour l'ensemble de la région Mid2 (Staller, P et al. 2001). Afin de caractériser l'interaction Mid2/Myc, la conservation de résidus sur Myc a été étudiée afin de trouver certains éléments déterminants de cette interaction. Les résidus V394 et K397 sur la surface externe de l'hélice II de Myc sont hautement conservés (pas dans Max), indiquant une possible importance dans son interaction avec Mid2. À partir de cette information, des mutants ont été effectués : V394A et K397E. Ces constructions diminuaient par 30 fois l'affinité entre Myc et Miz-1, suggérant que ces résidus sont situés sur le site de liaison de Myc pour Miz-1. La mutation V394D, quant à elle, abolissait totalement l'interaction Miz-1/Myc (Herold S, et al. 2002).

1.3.3 Compétition Miz-1/Max pour Myc

Il a été démontré que les gènes rapporteurs activés par Miz-1 peuvent être réprimés par une induction de l'expression de Myc par transfection d'un vecteur codant. De plus, cette répression par Myc peut être levée par une co-transfection avec un vecteur codant pour Max. Cela suggère que Miz-1 et Max compétitionnent pour l'interaction avec Myc (Peukert K, et al. 1997). Comme mentionné plus haut, les régions sur Miz-1 interagissant avec Myc ont été circonscrites (Mid1 et Mid2) par Peukert K, et al. Cependant, ce réseau d'interactions a été exploré pour une première fois par notre groupe de recherche (Bédard M, et al. 2016). À l'aide d'expériences d'immunoprécipitation, il a été démontré que les régions Mid1 et Mid2 interagissent avec le

bHLH-LZ de Myc – plus principalement le Mid2. Lorsque les mêmes expériences sont répétées, mais cette fois ci avec le mutant MycV394D (mutation abolissant l'interaction Miz-1/Myc, voir 1.3.2), il n'y a pas de Mids dans les immunoprécipitations – donc pas d'interaction détectée. Finalement, les mêmes expériences répétées avec le bHLH-LZ de Max et de Mad ne montrent aucune interaction, suggérant une haute spécificité des Mids pour Myc. De plus, la compétition Mid/Max pour Myc a été validée par les expérimentations d'immunoprécipitation effectuées. L'ajout de concentrations croissantes de bHLH-LZ de Max sur le complexe Mid2/Myc venait graduellement abolir l'apparition du complexe, montrant alors l'existence d'une compétition des constructions pour le bHLH-LZ de Myc (Bédard M, et al. 2016).

1.4 Protéines Intrinsèquement Désordonnées

1.4.1 Définitions et problématiques

Les protéines biologiquement actives mais sans structure ont longtemps été très peu étudiées dans la littérature de par leur nature difficilement discernable. Les premiers articles publiés sur de telles protéines, faute de termes et de pour les décrire tentaient tant bien que mal de les classer sous forme de protéines « étranges » ou « anormales ». Ce qui rend ces protéines dites anormales difficiles à étudier, c'est le fait qu'en utilisant les méthodes classiques d'études structurales de protéines (e.g. cristallographie à rayons X), l'information obtenue ne provient que d'une moyenne des conformations adoptées par la protéine. Les images superposées de toutes les structures adoptées par une protéine globulaire dans un environnement contrôlé offrent une image très claire et convergente, comme celles observées dans la PDB. Cependant, faire le même exercice en observant une protéine « anormale » offre une image démontrant un ensemble de structures extrêmement dynamique, témoignant d'une grande sensibilité de ces molécules pour les changements environnementaux et évoluant grandement dans le temps. Un nouveau champ d'études est donc mis au jour avec le terme *Intrinsically Disordered Protein* (IDP) (Uversky VN, 2016).

Les protéines intrinsèquement désordonnées (IDPs) sont de plus en plus reconnues comme ayant un rôle central dans plusieurs systèmes biologiques. Les IDPs (ainsi que les IDPRs – régions

intrinsèquement désordonnées) existent en tant qu'ensembles hétérogènes de structures fluctuant sous conditions physiologiques (e.g. dans une cellule) conférant un avantage pour plusieurs fonctions cellulaires. Cette fluctuation structurale en conditions physiologiques est due à une instabilité thermodynamique hautement sensible à différents stimuli tels que la liaison de la protéine à un ligand ou à son insertion dans la membrane plasmique, des changements dans l'environnement cellulaire et des modifications post-traductionnelles (PTMs) (Kriwacki RW, Chen J, 2018). Il est à noter que malgré les fluctuations structurales, les ensembles de structures adoptées par les IDPs en solution ne sont pas aléatoires dans leur forme non liée, et peuvent adopter des structures secondaires préformées ainsi que des régions désordonnées. Ces régions désordonnées possèdent des angles dièdres situés dans des zones moins favorables du diagramme de Ramachandran, et parfois conservent cette caractéristique suite à la liaison de l'IDP à un ligand, suggérant la conservation de désordre de certaines régions même après liaison (Fuxreiter M, 2018). Dans un système biologique d'ordre élevé, tel que l'Homme, les stimuli peuvent être multiples et même avoir des effets coopératifs. Les ensembles structuraux dynamiques que forment les IDPs sont les centres de régulation nécessaires à ce système de signalisation hautement complexe. L'ensemble structural dynamique peut être visualisé par la représentation de l'énergie libre inhérente à des IDPs (Fig.7). Ce type de protéines adopte plusieurs conformations de manière

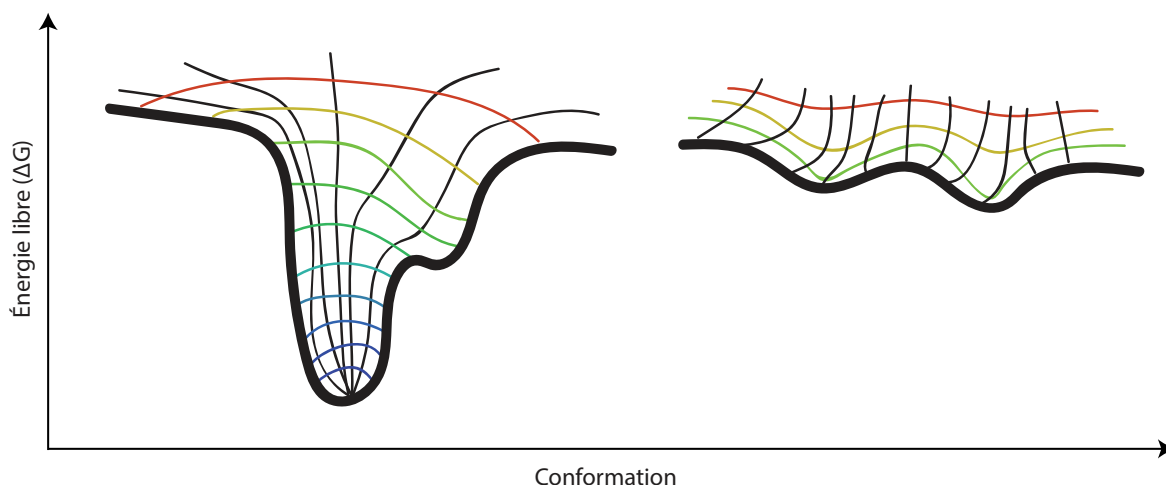


Fig.7 Énergie libre conformationnelle inhérente aux protéines globulaires et intrinsèquement désordonnées

Représentation de l'énergie libre en fonction des conformations adoptées par les protéines globulaires (à gauche) et par les protéines intrinsèquement désordonnées (à droite) lors du repliement. À noter que les protéines globulaires ont un minimum d'énergie libre beaucoup plus faible, et donc plus stable, que les IDPs.

Adapté de Burger V, et al. 2014

transitoire qui ne possèdent pas de minimum d'énergie libre stable. Ces protéines sont donc de manière inhérente en constant changement de conformation. De par leur nature centrale à la régulation des systèmes complexes, les IDPs sont fréquemment impliquées dans diverses pathologies humaines, principalement les maladies neurodégénératives, plusieurs types de cancers ainsi que le diabète (Kriwacki RW, Chen J, 2018).

Les protéines ont longtemps été classées selon une relation ordre-fonction (Stryer L, et al. Biochemistry 2002). Les IDPs viennent briser ce paradigme en ayant une relation désordre-fonction. Cependant, établir ces relations désordre-fonction pose un défi de taille puisque l'ensemble conformationnel adopté par les protéines intrinsèquement désordonnées est caractérisé incomplètement par les méthodes expérimentales disponibles pour le moment. Ces dites méthodes n'offrent que des paramètres structuraux moyens pour l'ensemble conformationnel – empêchant donc la mesure de caractéristiques structurales pour chacune des conformations précisément (Kriwacki RW, Chen J, 2018).

Une analyse des IDPs a révélé que leur absence de repliement est dû à une présence accrue de groupements chargés non-compensés ainsi qu'une quantité limitée d'acides aminés hydrophobes. En résumé, cela veut dire que les IDPs possèdent une charge nette élevée ainsi qu'une hydropathie moyenne basse – des éléments clés qui n'encouragent pas la formation d'une structure compacte en solution (Uversky VN, et al. 2000). De manière plus précise, les IDPs et les IDPRs sont pauvres en résidus tels que Cys, Trp, Tyr, Phe, Ile, Leu, Val et Asn. Elles sont également riches en résidus entraînant le désordre tels que Pro, Arg, Gly, Gln, Ser, Glu, Lys et Ala (Dunker A. et al. 2001). La conséquence de cette composition en acides aminés est une structure primaire beaucoup plus simple que pour la plupart des protéines ordonnées. Une structure primaire plus simple signifie qu'un volume plus faible d'information structurale peut être stocké – donc moins d'informations pour entraîner le repliement de la protéine. Les IDPs et les IDPRs doivent rencontrer un partenaire d'interaction pour obtenir l'information nécessaire au repliement (Uversky VN, 2013). Étant donné la versatilité structurale des IDPs, elles sont impliquées dans de nombreuses voies métaboliques de régulation, l'expression génique, la régulation du cycle cellulaire et de la signalisation. Cette régulation se fait via l'interaction avec des partenaires d'interaction tels que d'autres protéines, l'ADN, l'ARN ou de petits métabolites (Fuxreiter M,

2018). Ces partenaires d'interaction peuvent entrainer les IDPs dans différents modes d'interaction correspondant à différents rôles des IDPs/IDPRs (Fig.8) (Uversky VN, 2016).

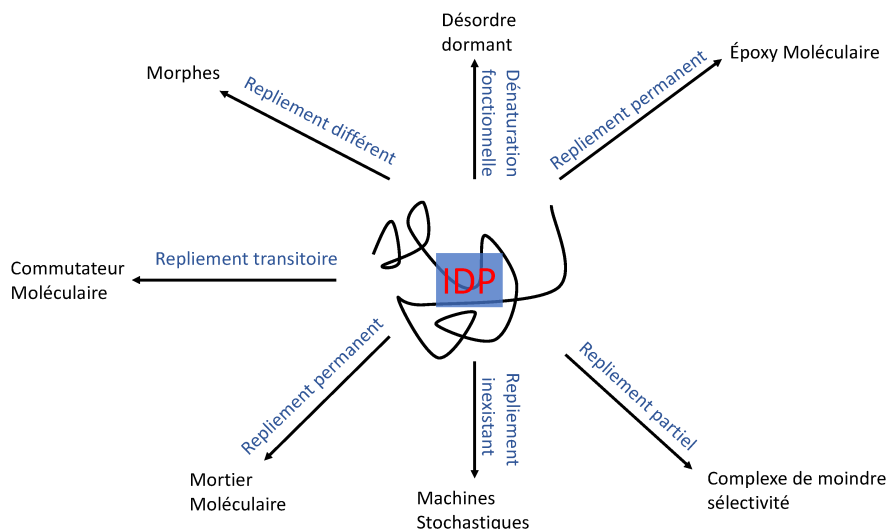


Fig.8 Modes de repliement des protéines intrinsèquement désordonnées

Domaines fonctionnels positionnés sur la structure primaire de c-Myc, de N à C terminal.

Adapté de Uversky VN, 2016.

1.4.2 Méthodes utilisées pour l'étude des IDPs

Il a été mentionné précédemment que les méthodes expérimentales pour étudier les IDPs sont inefficaces car elles fournissent des données moyennes de l'ensemble conformationnel adopté par ces protéines. Cette problématique peut être contournée par l'utilisation de nouvelles méthodes computationnelles, d'expérimentations de dispersion, telles que le *Small-Angle X-ray Scattering* (SAXS) ainsi que le *small-molecule Förster Resonance Energy Transfer* (smFRET). Ces méthodes fournissent des données sur la distribution des conformères des IDPs, au lieu de fournir des données moyennes de l'ensemble (Henriques J, et al. 2018)(Cragnell C, et al. 2018)(Zheng W, et al. 2018)(Holmstrom ED, et al. 2018). Le dichroïsme circulaire ainsi que la RMN à l'état liquide sont aussi des outils de choix pour l'élucidation des caractéristiques structurales qu'adoptent les IDPs lors d'interactions avec différents partenaires (Yu C, et al. 2016) (Ozenne V, et al. 2012). C'est d'ailleurs ces méthodes qui ont été préconisées pour caractériser Myc dans son interaction avec la région MID2 de Miz-1 dans ce mémoire.

1.4.3 La structure Poly-L-Proline II

Les protéines intrinsèquement désordonnées sont impliquées dans de nombreuses pathologies et problèmes de santé chez l'humain, et pour cette raison il serait intéressant de comprendre les mécanismes fonctionnels de ce type de protéine (Radivojac P, et al. 2007)(Ward JJ, et al. 2004). Il a été démontré à maintes reprises que les IDPs possèdent une propension conformationnelle intrinsèque pour la formation d'hélice Poly-L-Proline II (ci-après PPII). Cette structure se retrouve souvent ignorée dans la littérature et les logiciels de visualisation moléculaire (l'emphase est principalement mise sur les hélices- α et les feuillets- β), malgré son importance dans les processus biologiques. Cette structure a été reconnue comme étant fréquente dans les protéines globulaires, ayant même un ratio de conservation de 80 à 100% dans les familles de protéines ayant une similarité de séquence de 20% ou plus, ce qui approche les niveaux de conservation observés pour les hélices- α et les feuillets- β (Adzhubei AA et Sternberg MJ, 1994). Pour des raisons historiques, le nom Poly-L-Proline II est associé à ce type de structure malgré le fait que la plupart des PPII sont composés principalement de résidus qui ne correspondent pas à une proline, ou bien ne contiennent tout simplement aucune proline (Adzhubei AA et Sternberg MJ, 1993).

Les PPII sont des structures hélicales étendues de pas gauche et possèdent des angles dièdres moyens $\Phi = -75^\circ$ et $\Psi = +145^\circ$. Contrairement aux structures secondaires conventionnelles, les PPII ne possèdent pas de ponts hydrogènes internes stabilisants, puisque la structure hélicale est trop étendue. En comparaison, une hélice- α ($\Phi = -60^\circ$ et $\Psi = -50^\circ$) comporte 3.6 résidus totalisant une longueur de 5.4Å par tour alors qu'une structure PPII comporte 3 résidus totalisant une longueur de 9.3Å par tour (Fig.9). Cette conformation très étendue permet une excellente exposition des résidus au solvant, ce qui rend les peptides adoptant la structure PPII très propices à interagir avec différents partenaires (Chebrek R, et al. 2014) Malgré le fait que ces structures sont étudiées de manière insuffisante et peu répertoriées dans la PDB, il serait suggéré que les PPII ont un rôle fonctionnel important, spécialement dans les interactions protéine-protéine et la reconnaissance protéine-acide nucléique (Chevrier L, et al. 2013) (Ozenne V, et al. 2012).

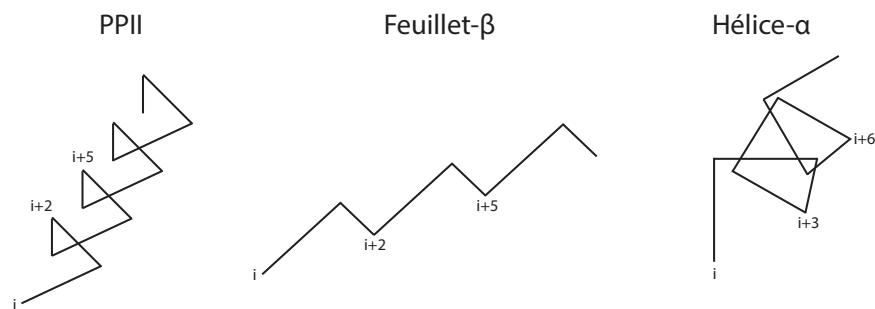


Fig.9 Comparaison des structures secondaire classiques avec PPII

Schématisation du tracé des C- α d'un polypeptide dans les différentes structures secondaires classiques ainsi que sous forme PPII.

Adapté de Adzhubei A, 2013.

La structure PPII serait un motif clé pour les protéines intrinsèquement désordonnées, spécialement pour de petits peptides. Le rayon hydrodynamique (R_h) des IDPs est modifié en fonction de la propension des différents acides aminés de leur structure primaire à former des structures PPII (Elam WA, et al. 2013). Une organisation globale de la structure des IDPs serait donc fortement dépendante de leur propension conformationnelle intrinsèque (Tomasso ME, et al. 2016). Le rayon hydrodynamique des IDPs est fortement influencé par des perturbations structurales localisées, telles que des substitutions d'acides aminés, des changements dans la charge nette, des réarrangements de charge et des changements de température (ce dernier n'étant pas local, mais vrai néanmoins). Le R_h des IDPs ayant une propension conformationnelle intrinsèque pour la structure PPII est plus large que le R_h pour une conformation en random coil d'un peptide de même longueur, ce qui suggère que les IDPs seraient favorisés à adopter une structure PPII (Tomasso ME, et al. 2016). Le R_h des IDPs peut même être modulé en fonction de leur teneur en acides aminés favorisant la formation de PPII, tels que Pro, Lys, Gln, Glu, Val et Ile et de l'état dénaturé des protéines. D'ailleurs, le R_h des protéines dénaturées est proportionnel à la proportion de résidus ayant une propension à former des hélices allongées PPII (Hamburger JB, et al. 2004) (Elam WA, et al. 2013).

1.4.4 Les IDPs et les transitions de phase liquide-liquide

Même si l'existence d'organelles sans membranes chez les eucaryotes et procaryotes a été suggérée à plusieurs reprises, ce n'est que tout récemment que l'on se penche sur leur utilité dans la cellule et ce qui provoque leur existence (Feric M, et al. 2016) (Aumiller WM et Keating CD, 2016). Les organelles sans membranes sont composées principalement de condensats de protéines, d'acides nucléiques, ou les deux combinés. Des exemples de telles organelles sont le nucléole, les corps de Cajal, les granules de stress, les corps P, etc. (Mitrea DM et Kriwacki RW, 2016). Un système (par exemple, le cytoplasme d'une cellule) où il y a présence de gouttelettes non miscibles aura une énergie libre plus faible (deux états non mélangés), et sera donc plus stable qu'un système où tout est homogène (un état mélangé). Le système où l'on retrouve deux états non mélangés s'appelle une transition de phase ou *liquid-liquid phase separation* (LLPS). Deux types d'interaction favorisent la formation de LLPS. D'abord, les interactions homotypiques entre les molécules d'une même phase et ensuite les interactions hétérotypiques entre les molécules de phases différentes (Gomes E et Shorter J, 2018). Un mélange dirigé par l'entropie est défavorisé dans de tels systèmes puisque les interactions homotypiques sont nettement plus fortes que les interactions hétérotypiques, favorisant alors la formation de deux états en séparation de phase possédant une énergie libre plus faible (Hyman AA, et al. 2014).

La cellule utilise judicieusement les avantages conférés par les séparations de phases. Non seulement ces phases peuvent être liquides, mais aussi sous forme d'hydrogels et même de solide, dépendamment de la force d'interaction des constituants de la séparation de phase. Les LLPS peuvent être utilisés comme principe dynamique d'organisation qui permettent à la cellule de compartimenter spatiotemporellement une biochimie spécifique, une infrastructure spécifique, ou les deux. De plus, ces compartiments cellulaires permettent le réarrangement interne et la diffusion de biomolécules à l'intérieur et à l'extérieur de la transition de phase (Boeynaems S, et al. 2018).

Bien que presque toutes les protéines peuvent effectuer des transitions de phases *in vitro*, seulement certaines protéines ayant des caractéristiques distinctes peuvent effectuer de telles transitions *in vivo*. Une caractéristique commune des protéines pouvant former des LLPS serait la multivalence de leurs domaines d'interaction. Bien que plusieurs protéines ordonnées présentent

des domaines d'interaction multivalents, les protéines intrinsèquement désordonnées sont fortement plus propices à posséder de tels domaines, ainsi que des structures primaires peu complexes (ayant beaucoup de répétitions) qui favorisent la formation de LLPS (Gomes E et Shorter J, 2018). Il a été démontré que les IDPs présentant ces caractéristiques pouvaient former favorablement des structures de type amyloïde (structures *cross- β*) qui constituent un type plus extrême de transition de phase (Shorter J et Lindquist S, 2005). Plusieurs types d'interactions permettent le maintien des transitions de phase. Ces interactions se doivent d'être faibles, puisque la concentration de plusieurs biomolécules dans une gouttelette a un coût énergétique. Les interactions faibles permettent de contrer le coût entropique pour la séparation de phase ainsi que le coût en énergie libre à l'interface pour créer une barrière entre les phases. Parmi ces interactions, on retrouve l'empilement π - π , les interactions cation- π , les interactions charge-charge et les contacts transitoires *cross- β* (Gomes E et Shorter J, 2018). En résumé, les protéines intrinsèquement désordonnées, de par leurs caractéristiques intrinsèques, ont tendance à former des séparations de phase liquide-liquide.

1.5 Objectifs de l'étude

Ce mémoire de maîtrise est une étude structurale et biophysique du complexe c-Myc/Mid2 ainsi que de sa compétition avec Max et de sa liaison à l'ADN. Dans cette optique, la région Mid2 a été séparée en trois peptides distincts afin de circonscrire la région minimale d'interaction de la région Mid2 avec c-Myc, ce qui nous permettrait au final d'avoir l'information structurale pour le développement d'un inhibiteur direct de c-Myc. Ils ont été séparés ainsi en suivant les résultats présentés par le groupe de recherche de Karen Peukert en 1997. En mutant certains résidus en C-terminal du Mid2, l'interaction Mid2/c-Myc était perdue (voir section 1.3.2). Les méthodes utilisées à ces fins sont expliquées en détails dans le chapitre 2. Les résultats présentés au chapitre 3 expliquent le raisonnement derrière le développement des différents peptides provenant de Mid2 (*Mid2-shorts*) et leur caractérisation biophysique. Un des points saillants du manuscrit est la découverte d'un peptide capable de séquestrer c-Myc dans une conformation stable thermodynamiquement qui compétitionne avec Max pour l'interaction. De plus, le manuscrit relate la caractérisation de la compétition entre Max et les Mid2s pour engager Myc et l'inhibition de la liaison à l'ADN de cette dernière.

2. Méthodes

Ce chapitre présente en détails les différentes méthodes utilisées afin d'obtenir les résultats présentés dans ce mémoire.

2.1 Obtention des constructions protéiques

2.1.1 Clonage

Plusieurs domaines de protéines et peptides ont été nécessaires afin de mener à bien ce projet de maitrise. Tout d'abord, les peptides Mid2s- $\alpha/\beta/\gamma$ ont été synthétisés sur phase solide par chimie Fmoc à la plateforme de synthèse peptidique de l'IPS. Les structures primaires des peptides synthétisés sont les suivantes:

Mid2s- α

GSHQGKAGIKILEPEEGSEVSVVTVDDMVTY

Mid2s- β

GSDDMVTLLATEALAATAVTQLTVVPVGAAVTADETY

Mid2s- γ

GSTADETEVLKAEISKAVKQVQEEDPNTHILY

Les domaines b-HLH-LZ de c-Myc et Max (c-Myc* et Max*) ont été bio-synthétisés à l'aide de la bactérie *E. coli*. La séquence d'ADN codant pour la structure primaire de c-Myc* :

5' GCC CCG CCA TAT GGT CAA ACG GCG CAC GCA TAA TGT ACT GGA
GAG ACA GCG GAG AAA CGA ATT GAA ACG GTC TTT CTT TGC GTT GCG
TGA CCA AAT CCC GGA GTT AGA AAA CAA CGA GAA AGC ACC GAA
AGT CGT CAT ATT AAA AAA AGC GAC GGC CTA CAT TCT TTC GGT ACA
AGC AGA AGA GCA AAA ATT GAT CTC CGA GGA AGA TCT TTT ACG

GAA GAG ACG CGA GCA ACT GAA GCA TAA ATT AGA GCA ACT GCG
TAA CTC CTA ATG AGG ATC CGC GCC 3'

a été obtenue de la compagnie IDT et clonée dans le *multiple cloning site* du plasmide pET3a (EMD-Millipore) à l'aide des enzymes de restriction NdeI et BamHI. Pour le clonage de Max*, les oligonucléotides suivants ont été utilisés :

5' CCG CCA TAT GGC TGA CAA ACG GGC TCA TCA TAA T 3'
5' GGC GGG ATC CTC ATT ACT CCA GTG CAC GGA CTT GCT GC 3'

Le clonage avec NdeI et BamHI a aussi été effectué dans le vecteur pET3A. La matrice était un plasmide contenant l'ADN codant pour p21 Max (Naud JF, et al. 2005) ayant quelques résidus de plus que désiré (la construction tronque p21 Max afin de ne conserver que le bHLH-LZ de Max, ci-après appelé Max*). Les plasmides ont été transformés dans des bactéries *E. coli* DH5 α compétentes (Transbiotech) afin d'amplifier le nombre de plasmides. Ensuite, ces plasmides ont été transformés dans des bactéries *E. coli* BL21(De3*) compétentes (New England Biolabs) afin de permettre la surexpression. La surexpression de protéines en bactéries compétentes s'effectue à l'aide d'une induction du système d'expression à l'aide d'IPTG. Cette molécule vient d'abord lever l'inhibition de l'opérateur de l'ARN polymérase du phage T7 en compétitionnant avec un répresseur, utilisant le même principe que l'opéron lactose. Un promoteur T7 est soigneusement placé en aval des séquences d'ADN codant voulues, ce qui permet à l'ARN polymérase du phage T7 de transcrire le tout. Cette polymérase transcrit 5 fois plus rapidement que les polymérases endogènes de la bactérie, ce qui permet d'avoir une forte surexpression. Un tel système se doit d'être inductible, sans quoi le stress sur la bactérie serait trop grand.

2.1.2 Surexpression de protéines

Les souches intégrées, obtenues par étalement sur pétri de milieu de culture LB-agar avec Ampicilline pour discrimination, ont été inoculées dans 1L de milieu de culture LB liquide afin d'obtenir une grande quantité de bactéries surexprimant c-Myc*. À une D.O.₅₉₅ de 0.6, les bactéries ont été induites avec 1mM d'IPTG. Après 12 heures d'induction à 30°C, les bactéries

sont culottées en centrifugeant à 8000g pendant 12 minutes. Le culot est resuspendu dans du tampon de lyse (50mM KH_2PO_4 , 500mM NaCl, pH4.5, cOmplete Roche 1X) à 3mL/g de culot. La lyse est complétée à l'aide d'une congélation à -80°C , et plusieurs cycles de congélation/décongélation rapide à l'azote liquide et à l'eau chaude. Suite à la lyse, 50 $\mu\text{g/mL}$ de DNase I est ajoutée, puis le tout est incubé 1 heure sous agitation à 37°C . 1mM de DTT est ajouté puis incubé à 37°C pendant 30 minutes sous agitation. Finalement, les bactéries lysées sont centrifugées à 19 000 rpm durant 1 heure. Pour la construction Max*, la protéine d'intérêt se retrouve dans le surnageant, et il est donc possible de poursuivre directement vers la purification. Pour la construction c-Myc*, la protéine d'intérêt se retrouve dans les corps d'inclusions, et doit donc être extraite. Deux constructions sont ainsi obtenues :

c-Myc*

**MVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVLKKATAYILSVQAEEOQLISE
EDLLRKRREQLKHKLEQLRNS**

Max*

**MADKRAHHNALERKRRDHKDSFHSLRDSVPSLQGEKASRAQILDKATEYIQYMRRKNHTHQQD
IDDLKRQNALLEQQVRALE**

Pour extraire c-Myc* des corps d'inclusions, le culot du lysat bactérien a été d'abord resuspendu dans 15mL de tampon de corps d'inclusion (0.5M GuHCl, 6M Urée, 100mM NaAc pH 5.0). 15mL de tampon de corps d'inclusion B (2M Urée) est alors ajouté aux corps d'inclusions resuspendus. Le tout est alors centrifugé à 19 000 rpm durant 1 heure à 4°C . La protéine c-Myc* se retrouve donc dans la partie soluble et est extraite des corps d'inclusions.

Dans le cas où la protéine d'intérêt doit être observée par RMN, elle doit être enrichie en isotopes non radioactifs ^{15}N et ^{13}C (voir 2.4.1). Pour se faire, le milieu LB mentionné ci-haut est substitué par un milieu minimal M9. Ce milieu de culture bactérien ne contient que les nutriments essentiels pour la croissance de bactéries, avec une source de carbone ^{13}C provenant de glucose et d'azote ^{15}N provenant de chlorure d'ammonium (8.50mM NaCl, 22.05mM KH_2PO_4 , 22.38mM

Na_2HPO_4 , 1mM MgSO_4 , 0.1mM CaCl_2 , 0.01mM FeSO_4 , 18.7mM $^{15}\text{NH}_4\text{Cl}$, 16.65mM ^{13}C -glucose).

2.1.3 Purification des protéines

Pour purifier les protéines du lysat bactérien, des colonnes SP-HP (GE Healthcare, Cation Exchange Columns 5mL) ont été utilisées à l'aide d'un appareil ÄKTAprime (Amersham Pharmacia Biotech) qui permet d'effectuer une FPLC (Fast Purification Liquid Chromatography). Les protéines dans le lysat ont été diluées 2-3 fois avec du tampon A1 (50mM NaAc pH 5.0) pour éviter d'avoir une concentration en sels trop élevée pour la liaison des protéines d'intérêt sur la colonne de chromatographie. Les colonnes ont d'abord été lavées avec 2 volumes de tampon A1, suivi d'un volume de lavage avec le tampon B1 (50mM NaAc pH 5.0, 5M NaCl) et de deux volumes de lavages avec le tampon A1. Le lysat bactérien dilué a alors été mis sur la colonne de chromatographie, à un débit de 3mL/min. Le volume ressortant a été conservé et observé par SDS-PAGE pour s'assurer que la protéine d'intérêt s'accrochait efficacement. Un volume de lavage a alors été mis sur colonne avec le tampon A1+10%B1, un volume de lavage avec le tampon A2 (50mM Citrate-Phosphate pH 3.0), et un volume de lavage avec le tampon A2+10%B2 (50mM Citrate-Phosphate pH 3.0, 3M NaCl). Le gradient d'élution s'est effectué de 10% à 60% de tampon B2 sur 100mL, récolté en fractions de 4-5mL à partir d'une conductivité de 85mS/cm². Après le gradient, un volume de colonne à 80% B2, et un autre à 100% B2 ont été mis sur colonne pour s'assurer de décrocher toute la protéine d'intérêt présente, ainsi que d'autres protéines contaminantes possiblement encore accrochées sur la colonne de chromatographie, pour s'assurer de bien laver la colonne. Les fractions récoltées lors de l'élution et les fractions de lavage ont été mises sur gel SDS-PAGE pour observer la séparation de la protéine d'intérêt avec les différentes protéines contaminantes possiblement présentes. Les fractions pures ont été combinées, et repurifiées sur colonnes d'exclusion de taille (GE Healthcare, Hitrap Desalting 5mL) afin de pouvoir choisir un tampon d'intérêt en fonction des expérimentations voulues, à l'aide d'un appareil ÄKTAprime (Amersham Pharmacia Biotech). Le volume présentant un pic d'absorbance à 280nm a été conservé, alors que le volume où la conductivité était supérieure à 2mS/cm² a été jeté (ce volume contient l'excès de sel dont on veut se débarrasser). Le volume contenant la protéine d'intérêt pure a été amené à 15% Acétonitrile, mis dans des tubes de centrifugeuse (*Falcon*) de

50mL équipés de couvercles troués à la seringue et congelés rapidement dans l'azote liquide. La solution obtenue a été congelée dans l'azote liquide et ensuite lyophilisée.

2.2 Dichroïsme circulaire

2.2.1 Principes inhérents à la méthode

Le dichroïsme circulaire exploite les propriétés physiques de l'interaction entre la lumière UV (180-350 nm) polarisée et les chromophores (liaison peptidiques, chaînes latérales aromatiques et ponts disulfures) organisées de façon asymétrique dans les protéines afin d'en extraire la composition en structures secondaires et la présence de structures tertiaires (ou quaternaire) stables thermodynamiquement. L'appareil utilisé pour cette méthode, le spectropolarimètre, génère de la lumière polarisée de façon linéaire à partir de composantes circulaires (de pas gauche et de pas droit, en phase). Lorsqu'elle interagit avec les structures asymétriques, une des composantes circulaires de cette lumière polarisée se retrouve absorbée différemment. Le signal dichroïque résulte de la différence entre les coefficients d'extinction molaire de la lumière polarisée de pas droit et de pas gauche ($\epsilon_R - \epsilon_L = \Delta\epsilon$). Cette différence mène à la déviation du plan de polarisation de la lumière incidente et le détecteur mesure, en millidegrés, cette déviation. Le signal CD est rapporté soit en millidegrés en ellipticité molaire ($[\theta] = \text{deg} \cdot \text{cm}^2/\text{dmol}$) (Greenfield NJ, 2006).

2.2.2 Utilisation du dichroïsme circulaire pour l'étude des protéines

Les protéines possèdent des éléments structuraux secondaires asymétriques, comme des hélices- α , des feuillets- β ou du *random coil*. Ces structures possèdent des spectres CD caractéristiques dans l'UV lointain (185-250nm), permettant une estimation rapide de structures secondaires de protéines. Ces structures sont estimées en déconvoluant à l'aide d'algorithmes disponibles sur internet, tel que Dichroweb, qui se basent sur différentes bases de données de spectres, et estiment le contenu en différentes structures secondaires présentes dans le spectre analysé (Whitmore L, Wallace BA, 2008). Il est possible d'observer la dénaturation de ces structures en suivant l'évolution du signal CD en fonction de la température. Il faut noter c'est

qu'en chauffant, nous faisons varier (augmenter) la constante d'équilibre ($K_U = [D]/[N]$) et les protéines perdent les structures fortement ordonnées qui leur donnent leur activité optique de même que leur structure tertiaire. Donc, la dénaturation entraîne la perte de signal dichroïque. Certaines protéines vont se dénaturer complètement et donner un signal dichroïque ressemblant à celui de petits polypeptides (e.g. collagène dénaturé), alors que d'autres ne se dénatureront que partiellement (*molten globule state*), gardant un signal dichroïque structural résiduel (Greenfield NJ, 2006). La plupart du temps, les expérimentations suivant l'évolution du signal s'effectuent à une longueur d'onde déterminée caractéristique de la structure secondaire observée. Par exemple, les spectres CD de protéines structurées majoritairement en hélices- α possèdent un minimum caractéristique à 222nm (Greenfield NJ, 2007).

L'information obtenue à l'aide de ces expérimentations est fortement pertinente, notamment pour connaître rapidement les caractéristiques physicochimiques et thermodynamiques d'une protéine. Parmi les paramètres pouvant être déterminés par dichroïsme circulaire, on retrouve l'enthalpie de van't Hoff ou enthalpie apparente de dénaturation (ΔH_{vH}), l'entropie de dénaturation (ΔS_U), le point de transition (T_M), l'énergie libre de Gibbs (ΔG_U°) de dénaturation et la population de protéine native/dénaturée. Cependant, le calcul de ces paramètres dépend du système étudié. Une protéine monomérique aura une population dénaturée suivant cette fonction :

$$P_U = \frac{K_U}{1 + K_U}$$

où K_U correspond à la constante d'équilibre de la transition entre la forme native et dénaturée :

$$K_U(T) = e^{(-\Delta G_U^\circ/RT)}$$

Le point de transition (T_M ou T_U° dans ce cas-ci), normalement obtenu expérimentalement par dichroïsme circulaire, correspond au point où l'énergie libre de Gibbs (ΔG_U°) correspond à 0 selon cette fonction :

$$\Delta G_U^\circ(T) = \Delta H_U^\circ(T_U^\circ) - T \left(\frac{\Delta H_U^\circ(T_U^\circ)}{T_U^\circ} \right) + \Delta C_{pU} \left((T - T_U) - \ln \left(\frac{T}{T_U^\circ} \right) \right)$$

Dans le cas d'un système binaire (e.g. un hétérodimère ou un homodimère), les précédentes fonctions ne peuvent pas s'appliquer. Le calcul de la population de protéine dénaturée doit maintenant inclure la constante de dissociation (K_D) du dimère et la concentration de protéines totales monomériques (pT) sous la forme de cette fonction :

$$P_U(T) = \frac{-K_D(T) + \sqrt{K_D(T)^2 + 8pTK_D(T)}}{4pT}$$

De la même manière qu'avec les équations pour un système monomérique, l'énergie libre de Gibbs (ΔG_U°) correspond à la même fonction présentée précédemment. Cependant, cette fois ci, un terme incluant la concentration de protéines totales monomériques (pT) a été ajouté pour l'adapter pour un système dimérique :

$$\Delta G_U^\circ(T) = \Delta H_U^\circ(T_U^\circ) - T \left(\frac{\Delta H_U^\circ(T_U^\circ)}{T_U^\circ} + R \ln(1.66pT) \right) + \Delta C_{pU} \left((T - T_U) - \ln \left(\frac{T}{T_U^\circ} \right) \right)$$

À l'aide des équations présentées dans cette section, il est possible de calculer de nombreuses caractéristiques physicochimiques inhérentes au système étudié (monomères, dimères) (Naud JF, et al. 2003). Les expériences de dénaturation nous permettent aussi d'observer la présence d'intermédiaires de repliement, nécessitant l'utilisation d'autres fonctions qu'il ne serait pas pertinent de présenter ici. Le système Mid2s- $\alpha/\beta/\gamma$ - c-Myc est grossièrement approximé en hétérodimère, faute de meilleures informations à cet égard.

La prise de spectres pendant la dénaturation (thermique ou chimique) peut aussi venir fournir des informations fortement pertinentes, notamment la présence d'un point isodichroïque. Le point isodichroïque correspond à la longueur d'onde où le signal CD n'évoluera pas en fonction de la température. De plus, l'évolution du signal de part et d'autre du point isodichroïque nous informe sur l'équilibre entre différentes structures via des spectres caractéristiques. Par exemple, un équilibre entre une structure Poly-L-Proline-II (voir section 1.4.3) et *random coil* aura un point isodichroïque à 209nm, incluant une perte de signal entre 185-209nm et un gain de signal entre 209 et 250nm en fonction de la température (Paz A, et al. 2008). Un équilibre entre une structure en hélice- α et *random coil* aura un point isodichroïque à 203nm, incluant un gain de signal entre 185-203nm et une perte de signal entre 203-250nm en fonction de la température (Paz A, et al. 2008).

2.2.3 Protocole d'utilisation du dichroïsme circulaire

Tout d'abord, les protéines d'intérêt purifiées et lyophilisées (voir 2.1) ont été solubilisées dans le tampon de dichroïsme circulaire (10mM NaCl, 10mM Na₂HPO₄, pH5.5). La faible quantité

de sel dans le tampon permet d'éviter d'avoir un signal CD saturé par l'absorption des UV lointain. Le spectre du tampon a été soustrait de la contribution spectrale totale. Les peptides Mid2s- $\alpha/\beta/\gamma$, c-Myc* et Max* ont été étudiés en CD à des concentrations variant entre 10 μ M et 100 μ M dépendant des ratios de titration utilisés. Pour les expériences contenant de l'ADN, la concentration d'ADN était de 20 μ M double-brin, la concentration de c-Myc* et de Max* était de 20 μ M et la concentration de peptides Mid2s- $\alpha/\beta/\gamma$ variait entre 20 μ M 100 μ M. Les oligos provenaient d'IDT, de séquence [5'-GAG TAG CAC GTG CTA CTC-3' et 5'-GAG TAG GGA TCC CTA CTC-3'] pour le Ebox canonique et l'ADN non-spécifique respectivement. Les oligos ont été resuspendus dans l'eau et hybridés en dénaturant en chauffant à 95°C pour 5 minutes et en laissant refroidir à température ambiante pour 30 minutes. La majorité des expérimentations ont été effectuées à des concentration de protéines de 50 μ M. Le volume utilisé dans les cuvettes de dichroïsme circulaire (Quartz Suprasil, 1mm) était de 160 μ L. Le spectropolarimètre utilisé était le Jasco J-810 couplé à un thermostat de type Jasco-Peltier, purgé de manière constante à l'azote gazeux. Les expérimentations de dénaturation thermiques ont été effectuées à 222nm ou à 210nm, de 5°C à 95°C à 1°C/min.

2.3 Electrophoretic Mobility Shift Assays (Essais de retardement sur gel)

2.3.1 Principes inhérents à la méthode

Les expérimentations de retard sur gel d'électrophorèse (EMSAs) sont couramment utilisées pour l'étude d'interactions protéines-acides nucléiques. Un gel de polyacrylamide natif est utilisé. Le puit contrôle contient l'ADN fluorescent d'intérêt sans protéines – il y aura donc l'apparition d'une seule bande correspondant à l'ADN non lié lors de la révélation. Si une protéine d'intérêt est capable de lier l'ADN, une bande correspondant au complexe protéine-ADN de plus grande taille et de mobilité moindre apparaîtra. De cette manière, il est possible de suivre l'interaction ou non de protéines avec des acides nucléiques (Chen R, 2011) (Maltais L, et al. 2017)

2.3.2 Protocole d'utilisation des EMSAs

Les gels d'électrophorèse natifs étaient à 6% polyacrylamide, au Tris-acétate à pH 8.0. L'homodimère Max*/Max* ou l'hétérodimère c-Myc*/Max* a d'abord été incubé 30 minutes à température ambiante à des concentrations de 1000nM pour l'homodimère Max*/Max* ou 500nM/500nM pour l'hétérodimère c-Myc*/Max*. Ensuite, les peptides Mid2s- $\alpha/\beta/\gamma$ ont été ajoutés à des concentrations variant de 500nM à 100 000nM dépendamment de la titration, et incubés 30 minutes à 37°C. L'ADN fluorescent (couplé à du iFluor) de séquence [5'-GAG iFluorTAG CAC GTG CTA CTC-3' et 5'-GAG iFluorTAG GGA TCC CTA CTC-3'] pour le Ebox canonique et l'ADN non spécifique respectivement, ont été resuspendus dans du TE-NaCl pH 8.0. Les oligos ont été pré-équilibrés 5 minutes avant leur ajout aux protéines dans le tampon de liaison 2X (Tris-HCl 40mM pH 8.0, KCl 150mM, BSA 50 μ g/mL, glycérol 10%). L'ADN est alors ajouté aux protéines, pour arriver à un volume final de 20 μ L (10 μ L de pré-mix ADN et 10 μ L de pré-mix protéines), ce qui nous donne le tampon de liaison à une concentration de 1X (Tris-HCl 20mM pH 8.0, KCl 75mM, BSA 25 μ g/mL, glycérol 5%). Le mélange ADN-protéines est incubé 5 minutes à température ambiante, et 4 μ L de ce mélange est ajouté dans un puits. Le gel d'électrophorèse natif a préalablement été migré à vide à 100V pendant 30 minutes. Une fois les puits remplis, le gel natif est migré durant 40 minutes à 100V. Le gel est alors révélé pour un temps d'exposition de 2.0 secondes à l'aide d'un système Molecular Imager VersaDoc MP4000.

2.4 Résonance Magnétique Nucléaire à l'état liquide

La spectroscopie RMN en solution est un outil puissant pour l'analyse de la structure et la dynamique des protéines et de leurs complexes. Ce type de spectroscopie repose sur les propriétés magnétiques du noyau atomique. Lorsque placés dans un champ magnétique externe, certains noyaux (qui ont un nombre de spin $I = \frac{1}{2}$) résonnent (basculent d'un état de basse énergie, aligné avec le champ externe à un état de haute énergie, aligné contre le champ externe) à des fréquences radio caractéristiques (Dans un champ de 14.1T : ^1H – 600MHz, ^{13}C – 150MHz, ^{15}N – 60MHz) du spectre électromagnétique. Ces fréquences de résonance renseignent sur la structure moléculaire (ou des changements de structure) dans laquelle les atomes se situent (suivant la formation de

complexes). Cette section portera sur l'utilisation de ces informations pour l'étude de la formation de complexes protéine/protéine ou protéine/ligand (Jacobsen NE, 2007).

2.4.1 Principes inhérents à la méthode

Les isotopes stables utilisés en RMN biologiques comme le ^{13}C et le ^{15}N doivent être enrichis dans la molécule d'intérêt afin d'obtenir un signal assez fort – leur abondance naturelle étant pour la plupart insuffisante (Fig.10). L'enrichissement des molécules en isotopes (spécialement ^{15}N et ^{13}C) s'effectue comme décrit en 2.1.2.

Noyau	Abondance (%)	Sensibilité	Fréquence (MHz)
^1H	100	1.0	600
^{13}C	1.1	0.016	150
^{15}N	0.37	0.001	60
^{31}P	100	0.066	243

Fig.10 Fréquences de résonance de noyaux importants pour la spectroscopie RMN

Fréquence de résonance de noyaux soumis à un champ magnétique de spectromètre RMN conventionnel (Bruker Avance III – 600MHz)
Adapté de Jacobsen NE, 2007.

La fréquence de résonance n'est pas seulement une caractéristique du type de noyau, mais varie aussi en fonction de la position de l'atome à l'intérieur d'une molécule – c'est ce qu'on appelle l'environnement chimique. Cette variation survient puisque les électrons impliqués dans les liaisons interatomiques créent leur propre champ magnétique, qui s'oppose généralement au champ externe, ce qui vient influencer le champ magnétique net ressenti par les noyaux. Ce champ magnétique externe est influencé principalement par la proximité d'atomes électronégatifs (O, N, Cl, etc.) dans le réseau de liaisons, ainsi que la proximité de groupes insaturés ($\text{C}=\text{C}$, $\text{C}=\text{O}$, Aromatiques) à travers l'espace. Cette variation infime du champ magnétique, appelée déplacement chimique (δ), est mesurée en ppm. La fréquence de résonance d'un atome dans une molécule doit donc être adaptée au déplacement chimique selon cette fonction :

$$\text{Fréquence de résonance} = \nu(1.0 + \delta \times 10^{-6})$$

Lorsqu'on corrèle l'intensité de l'énergie de transition avec la fréquence de résonance obtenue ici, nous obtenons un spectre 1D ^1H -NMR (Jacobsen NE, 2007).

2.4.2 Spectres 1D ^1H -NMR

Les spectres 1D ^1H -NMR montrent des signaux pour chacun des atomes d'hydrogène (protons) de la molécule. Les déplacements chimiques en proton se retrouveront entre 0 et 13 ppm (autour de la fréquence de résonance de 600MHz du proton, dans le cas d'un spectromètre RMN de 14T). Le déplacement chimique des protons dans un environnement contenant des « hydrocarbures » sans atomes électronégatifs ou de groupements insaturés, se retrouve proche de l'extrémité droite du spectre (*upfield*) comme suit : $\text{CH} > \text{CH}_2 > \text{CH}_3$ (1.6 ppm, 1.2 ppm, 0.8 ppm). Chaque substitution ramène le déplacement chimique vers la gauche (*downfield*). La proximité d'un proton d'une insaturation va affecter fortement son déplacement chimique vers la gauche, puisque l'atome se retrouve dans le plan de l'insaturation. Par exemple, un méthyle (CH_3) attaché à un aldéhyde ($\text{C}=\text{O}$) aura un déplacement chimique de 2.1 ppm plutôt que de 0.8 ppm. La combinaison entre l'électronégativité et l'insaturation peut être observée avec le déplacement

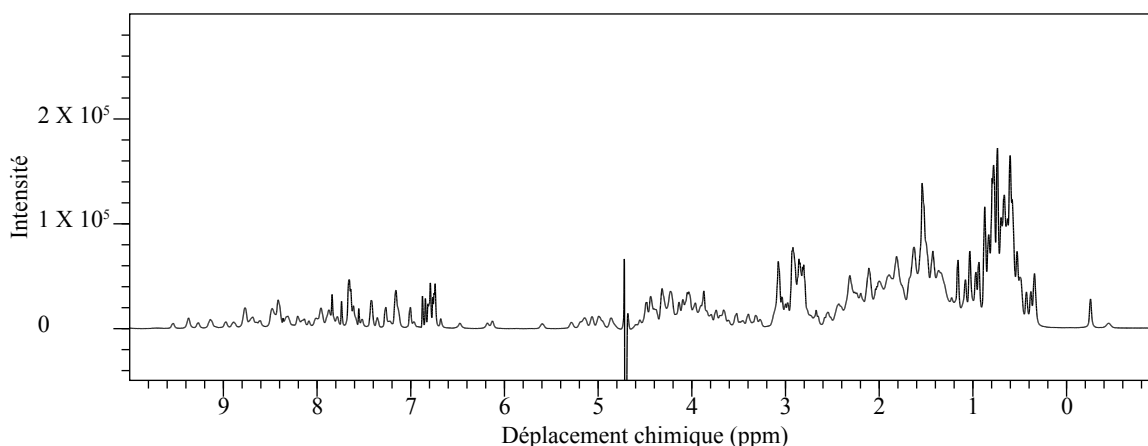


Fig.11 ^1H -NMR de l'ubiquitine UbKEKS

^1H -NMR de l'ubiquitine UbKEKS associé au ^1H - ^{15}N -HSQC de la figure 11

Figure fournie par Patrick Delattre, Laboratoire Pierre Lavigne

chimique du proton sur un groupement aldéhyde ($\text{H}-\text{C}=\text{O}$) qui possède un déplacement de 10 ppm. Les protons situés sur des cycles aromatiques se situent entre 7-8 ppm, et les protons situés sur des alcènes entre 5-6 ppm (Jacobsen NE, 2007). Il est aussi important de comprendre les déplacements chimiques associés aux amides, nécessaires à la compréhension de la prochaine section. D'abord,

les amines primaires auront un déplacement chimique entre 1-2.5 ppm, ensuite, les amines attachées à un cycle aromatiques auront un déplacement chimique entre 1.5-5 ppm, et finalement, les amides auront un déplacement chimique entre 5-10 ppm. Le déplacement chimique des amides est d'une importance capitale pour la compréhension des spectres ^1H - ^{15}N -HSQC, puisque les signaux observés ne se retrouveront que dans cette région du spectre en proton, étant donné que l'on observe les résonances associées aux amides. La figure 11 est un spectre 1D ^1H -NMR de l'ubiquitine UbKEKS qui est associé au spectre ^1H - ^{15}N -HSQC présenté en 2.4.3. Notez que seulement les résonances dans l'échelle des amides possèdent un pic croisé.

2.4.3 Spectres 2D ^1H - ^{15}N -HSQC

Le spectre 2D ^1H - ^{15}N -HSQC (Heteronuclear Single-Quantum Coherence) nous offre un signal pour chaque groupe ^1H - ^{15}N lié de manière covalente (Bodenhausen G, Ruben DJ, 1980). Ce spectre consiste en la corrélation du déplacement chimique en proton (^1H) comme présenté en 2.4.2 avec le déplacement chimique en amide (^{15}N). Le spectre obtenu s'observe comme une carte topographique. Pour une protéine dans des conditions optimales, le spectre ^1H - ^{15}N -HSQC contiendra un pic (Note : pour simplification, le terme pic croisé (*crosspeak*), représentant la

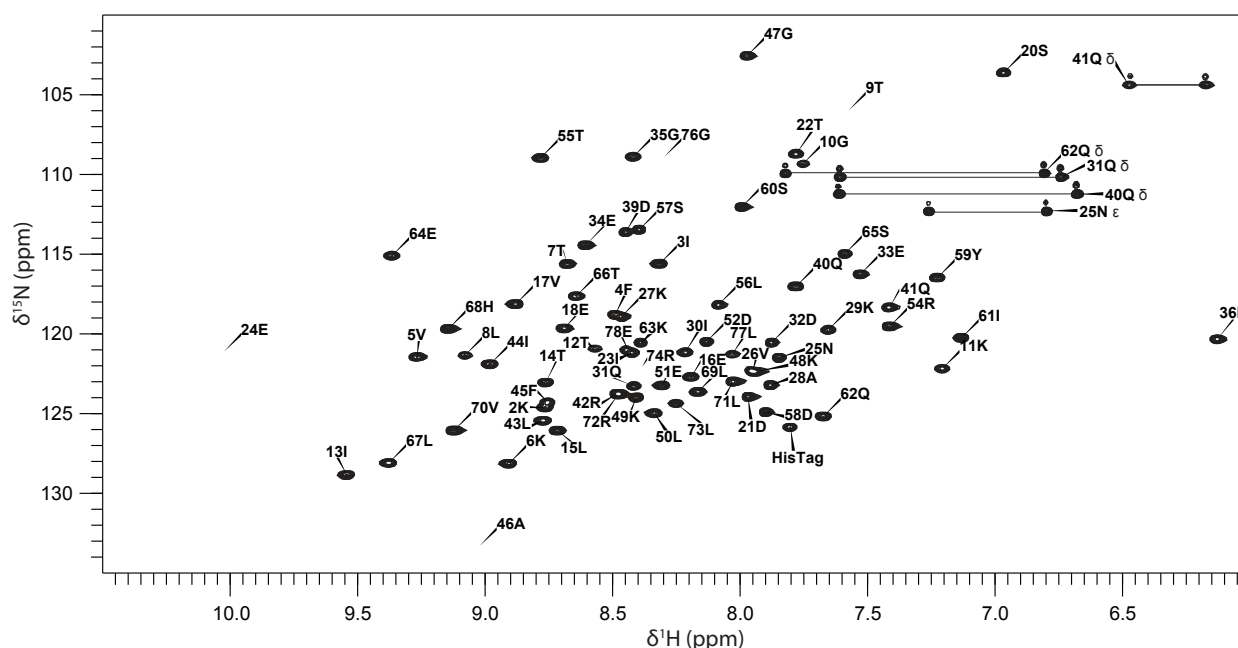


Fig.12 ^1H - ^{15}N -HSQC de l'ubiquitine UbKEKS assignée
Exemple de ^1H - ^{15}N -HSQC de protéine globulaire bien repliée, UbKEKS
Figure fournie par Patrick Delattre, Laboratoire Pierre Lavigne

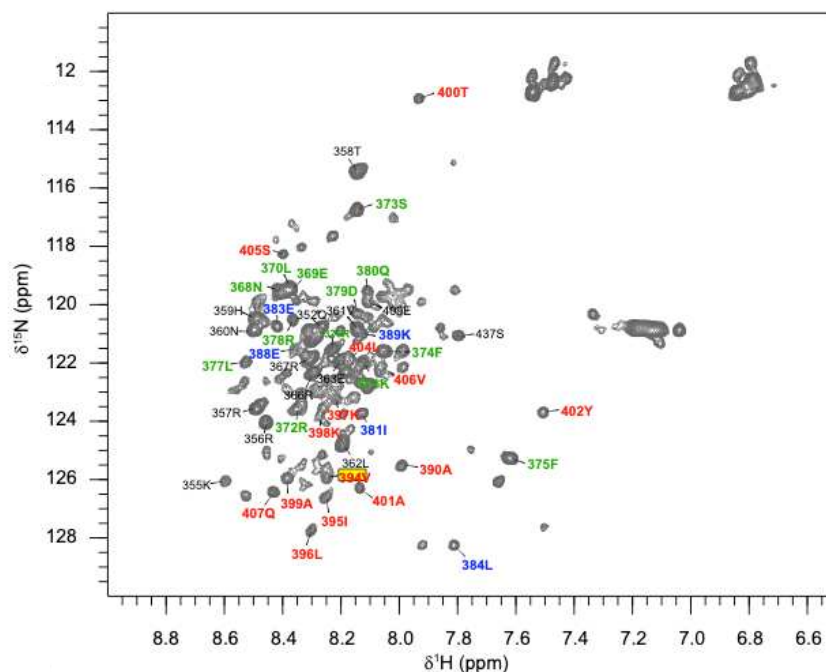


Figure fournie par Danny Létourneau, Laboratoire Pierre Lavigne

2.4.4 Titrations par ^1H - ^{15}N -HSQC

La spectroscopie RMN a un rôle de plus en plus déterminant dans le processus d'identification de molécules potentielles pour le développement de drogues, par sa capacité de reconnaître et caractériser rapidement la liaison d'un ligand en faisant fi de la fonction de la protéine cible (Hajduk et Greer, 2007). L'outil de base utilisé dans ces essais est le ^1H - ^{15}N -HSQC, puisqu'il permet de suivre l'évolution du déplacement chimique de résidus selon leur environnement chimique. Donc, si un résidu de la protéine d'intérêt est affecté par un ligand (ou une protéine interactante) le déplacement chimique sera modifié, et nous permettra d'identifier les résidus impliqués dans l'interaction (Ziarek JJ, et al. 2011). Il est à noter qu'afin d'avoir l'identité des résidus dans le ^1H - ^{15}N -HSQC, ainsi que l'information structurale du complexe protéine-ligand, plusieurs expérimentations RMN supplémentaires sont nécessaires (e.g. HNCACB, CBCA(CO)NH, CC(CO)NH, HNCO, etc.). C'est ce qui a été effectué afin d'avoir l'assignation de l'ubiquitine UbKEKS à la Fig.12. Mais lorsqu'une protéine est déjà assignée, la rapidité de la titration de la protéine d'intérêt avec un ligand via le ^1H - ^{15}N -HSQC est un atout considérable.

Afin de comprendre les types d'échange que les résidus peuvent subir lors d'une interaction avec un ligand, il faut d'abord comprendre la dynamique d'association et de dissociation. Dans le cas d'une protéine P liant de manière réversible un ligand L, l'équilibre ($\text{P} + \text{L} \leftrightarrow \text{PL}$) est caractérisé par des constantes de réaction k_{on} et de réaction inverse k_{off} . La constante de dissociation K_d équivaut à :

$$K_d = \frac{[\text{P}][\text{L}]}{[\text{PL}]} = \frac{k_{\text{on}}}{k_{\text{off}}}$$

Le K_d correspond à la concentration de ligand et de protéine nécessaires pour saturer la moitié des sites de liaison. Il est possible de connaître la vitesse de réaction et de réaction inverse en calculant $[\text{P}][\text{L}]k_{\text{on}}$ et $[\text{PL}]k_{\text{off}}$ respectivement. (Williamson MP, 2013).

Lorsqu'on effectue une titration par ^1H - ^{15}N -HSQC d'une protéine avec ligand, la vitesse de réaction dictée par k_{on} et k_{off} va influencer les effets observés sur les déplacements chimiques des pics. Par exemple, lorsque l'échange est lent dans l'échelle de temps du déplacement chimique (e.g. μs , ms), c'est que le k_{off} est significativement plus petit que la différence en Hz entre le

déplacement chimique de la protéine liée et libre. Cela résulte alors en une disparition graduelle du signal de la protéine libre associé à l'apparition de signal graduelle du signal de la protéine liée dans le ^1H - ^{15}N -HSQC (Fig. 14). L'intensité des signaux observés est proportionnelle à la concentration de la population de protéine libre et liée. Lorsque le k_{off} est significativement plus élevé que la différence de déplacement chimique en Hz, le signal de la protéine libre va se déplacer simplement vers le signal de la protéine liée, correspondant à chaque point de la titration au déplacement chimique moyen des populations de protéine liée et libre (*population weighted average*) comme présenté dans la Fig.14. Un problème survient lorsque l'échange est moyen dans l'échelle de temps du déplacement chimique (lorsque le k_{off} est similaire à la différence entre le déplacement de la protéine libre et de la protéine liée). Le signal se retrouve élargi (*broadened*) et se déplace à la fois. Il est donc difficile, voire impossible, de mener une titration à l'aide de ^1H - ^{15}N -HSQCs lorsque l'échange est moyen, puisque les signaux importants pour l'analyse disparaissent (Williamson MP, 2013).

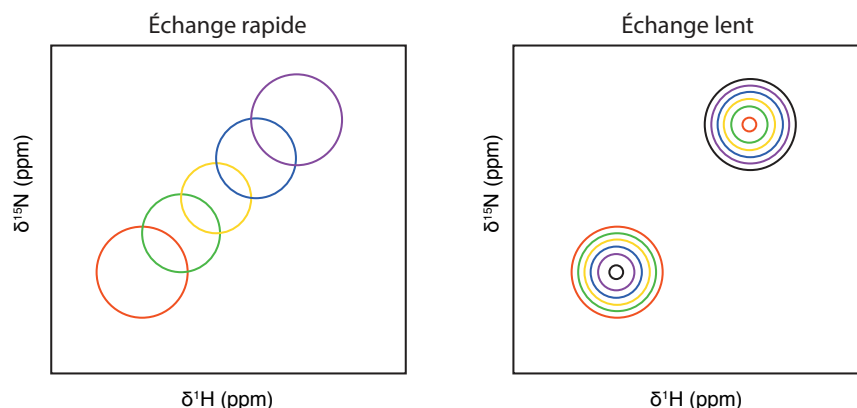


Fig.14 Effets des types d'échanges sur les pics en ^1H - ^{15}N -HSQC
Effet de l'échange rapide et de l'échange lent sur un pic de spectre ^1H - ^{15}N -HSQC
Adapté de Williamson MP, 2013

2.4.5 Protocole expérimental des titrations par ^1H - ^{15}N -HSQC

Les protéines d'intérêt ont été purifiées et lyophilisées comme mentionné en 2.1, avec des isotopes non-radioactifs ^{15}N ou $^{15}\text{N}+^{13}\text{C}$ dépendamment des expérimentations RMN voulues. Il est coûteux de produire de la protéine marquée $^{15}\text{N}+^{13}\text{C}$ si l'objectif est seulement d'obtenir de

l'information provenant de ^1H - ^{15}N -HSQCs. Il est préférable d'utiliser ces protéines doublement marquées lorsque l'on désire assigner les résidus du ^1H - ^{15}N -HSQC et obtenir de l'information structurale sur la protéine d'intérêt.

La protéine d'intérêt lyophilisée est resuspendue dans le tampon RMN (50mM NaCl, 50mM KH_2PO_4 pH5.5, 0.02% NaN_3 , 10% D_2O , inhibiteurs de protéases cOmplete EDTA de Roche 1X) à une concentration de 50 μM . Les ^1H - ^{15}N -HSQCs sont acquis durant 18 heures (448 scans, FID à 128), à une largeur spectrale de 10 000 Hz en proton (16.67 ppm) et de 2200 Hz en amide (36.67 ppm). La titration a été effectuée par incréments de 50 μM de ligand jusqu'à une concentration de 500 μM . Les spectres obtenus ont été traités avec le programme TopSpin de Bruker en appliquant une fonction d'apodization *sine bell* déplacé de 90 degrés, et observés à l'aide de CCPNMR (Vranken WF, et al. 2005).

3. Résultats

3.1 Séquestration de c-Myc par les peptides Mid2s

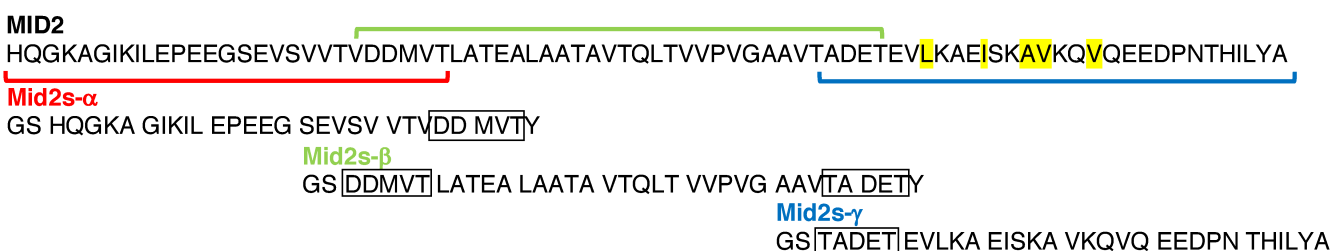
3.1.1 Les peptides Mid2s- α , β et γ adoptent une conformation polyproline-II

Le groupe de recherche de Martin Eilers a établi en 1997 des déterminants structuraux importants pour l'interaction c-Myc/Miz-1. Leurs expérimentations montraient que les régions Mid1 et Mid2 sur Miz-1 étaient nécessaires pour observer une interaction entre les deux facteurs de transcription (Peukert et al. 1997). Plus précisément, il a été démontré que des mutations de résidus d'acides aminés avec une propension élevée à la formation d'hélice- α dans la région Mid2 par des prolines diminuaient fortement ou abrogeaient l'interaction entre Miz-1 et c-Myc (voir 1.3.2). Depuis, le paradigme dicte que la région Mid2 adopte une structure d'hélice- α amphipatique, et que les résidus importants pour la formation du complexe sont situés en C-terminal de ladite région. Afin de circonscrire le Mid2 minimal, ou d'identifier les déterminants essentiels, nous avons séparés la région en trois peptides : Mid2s- α ; a.a. 641-664, Mid2s- β ; a.a. 660-692, and Mid2s- γ ; a.a. 688-715 (voir Figure 15A). Le peptide Mid2s- γ contiendrait, selon la littérature, les déterminants structuraux nécessaires à l'interaction Mid2/c-Myc. Les peptides Mid2s- α et Mid2s- β sont de taille similaire et font office de contrôle pour des interactions non-spécifiques ou d'autres déterminants structuraux nécessaires à l'interaction.

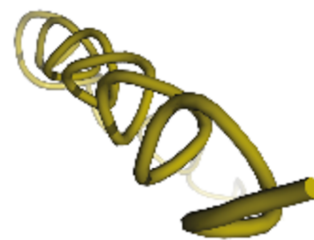
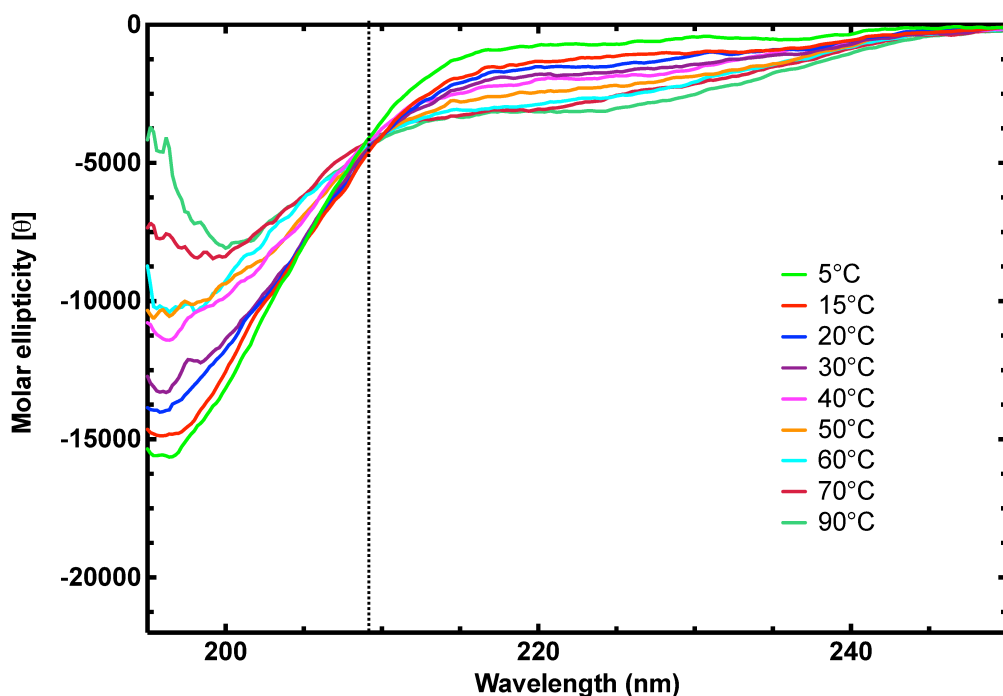
Il a été récemment démontré que la région Mid2 entière, selon sa signature spectrale de dichroïsme circulaire, n'adopte pas de structure en hélice- α comme prédit (Bédard M, et al. 2016). Comme on pourrait s'y attendre, les résultats présentés dans ce mémoire montrent que les trois peptides provenant de la région Mid2 n'adoptent pas de structure en hélice- α lorsqu'étudié indépendamment en dichroïsme circulaire. Le fait que les trois peptides ne possèdent pas les minima caractéristiques d'une structure en hélice- α à 208 et 222nm vient supporter cette hypothèse (voir Figures 15B-D-E). Il est important de noter qu'un équilibre structural entre hélice- α et *random coil* aurait, en signal dichroïque, un point isodichroïque à 203nm en fonction de la température. Cependant, les peptides provenant de Mid2 possèdent un tel point à 209nm, ce qui

est indicateur d'un équilibre structural entre Poly-L-Proline II et *random coil* (Fig. 15C) (voir 2.2.2) (Paz A, et al. 2008). Il serait possible d'argumenter que la structure Poly-L-Proline II nécessite une haute teneur en résidus proline pour être adoptée. Cependant, il a été démontré que cette structure ne nécessite pas nécessairement des résidus proline pour être adoptée, et même qu'elle est souvent caractéristique de protéines intrinsèquement désordonnées (Woody RW, 2009) (Voir 1.4.3). En effet, la forme dénaturée des protéines ainsi que les IDPs en contiendrait une large proportion (Elam W, et al. 2013)(Adzhubei AA, et al. 2013)

A



B



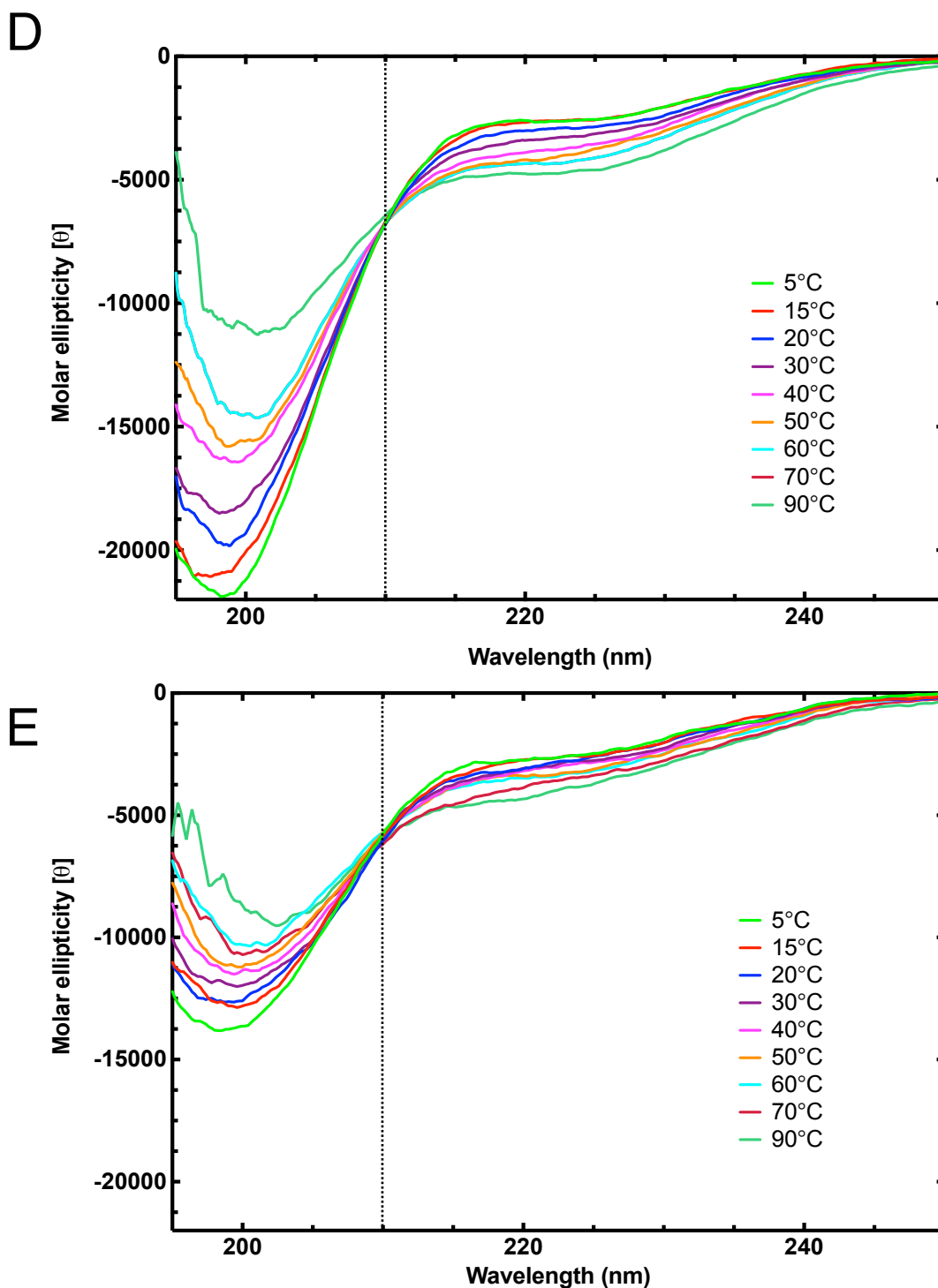


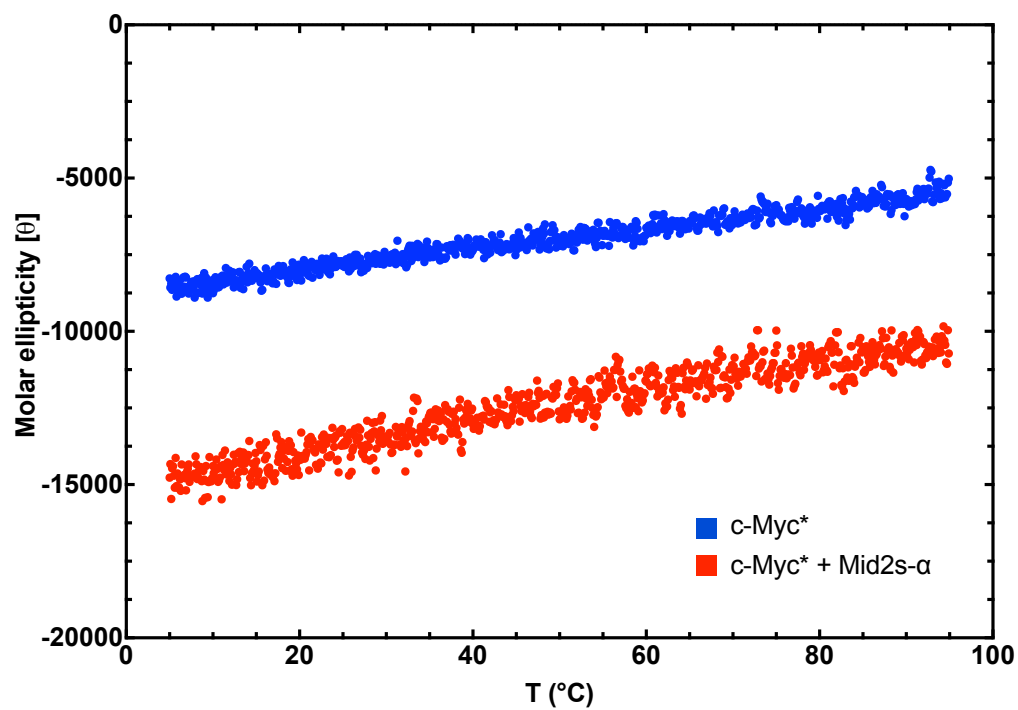
Fig.15 Caractérisation des peptides Mid2s- α / β / γ par dichroïsme circulaire

(A) Structure primaire du Mid2 complet. Les peptides Mid2s- α / β / γ sont isolés dans les accolades en rouge, vert et bleu respectivement. Les régions se chevauchant sont encadrés, les sites de mutation de Peukert (1997) sont surlignés en jaune. (C) Conformation Poly-L-Proline II (B-D-E) Spectres CD dans l'UV lointain des constructions Mid2s à températures croissantes, montrant des points isodichroïques (B) Mid2s- α à 209nm (D) Mid2s- β à 210nm (E) Mid2s- γ à 210nm

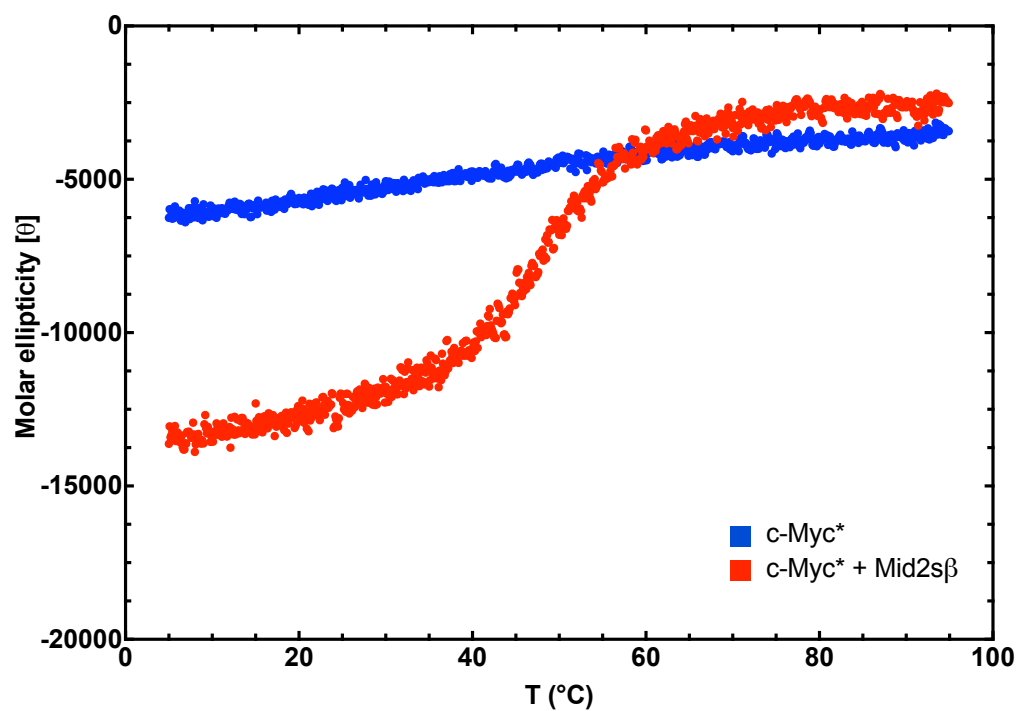
3.1.2 Mid2s- β forme un complexe spécifique avec c-Myc plus stable que l'hétérodimère c-Myc/Max

Bien que les trois peptides provenant de la région Mid2 sur Miz-1 aient été caractérisés, il faut déterminer leur interaction ou non avec le bHLH-LZ de c-Myc. Pour ce faire, des expérimentations de dichroïsme circulaire en fonction de la température ont été effectuées à un ratio 1:1 Mid2s/c-Myc à la longueur d'onde du point isodichroïque (déterminé en 3.1). L'observation du signal dichroïque nous permet de retirer la contribution des peptides Mid2s au signal, et ainsi nous permet d'observer l'évolution du signal pour c-Myc seulement. En d'autres mots, l'apparition d'une transition définie dans le signal dichroïque à cette longueur d'onde serait significative de la formation d'un complexe stable thermodynamiquement par c-Myc. De prime abord, le peptide Mid2s- γ ne semble pas affecter la stabilité thermodynamique du bHLH-LZ de c-Myc, puisqu'aucune transition coopérative n'apparaît dans le mélange expérimental 1:1 (voir Fig.16A). Le peptide Mid2s- α , quant à lui, induit une légère inflexion dans la courbe de dénaturation de c-Myc au point isodichroïque (voir Fig.16C), signifiant que ce peptide contient des résidus capables d'interagir avec le bHLH-LZ de c-Myc. Cependant, ce peptide possède une affinité apparente plus faible pour c-Myc que celle du peptide Mid2s- β . Le peptide Mid2s- β induit l'apparition d'une transition coopérative dans la dénaturation thermique de c-Myc au point isodichroïque, avec un T° de 48°C (voir Fig.16B). La simulation de cette courbe de dénaturation avec un modèle de dissociation d'un hétérodimère nous donne un $K_D(25^\circ\text{C})$ de $4.4 \cdot 10^{-7}$ et un $K_D(37^\circ\text{C})$ de $7.1 \cdot 10^{-6}$. Les K_D apparents du complexe à 25°C et 37°C sont de $1.1 \cdot 10^{-4}$ et $1.9 \cdot 10^{-3}$ respectivement, avec des populations correspondantes à 0.43 et 0.09 (voir Fig.16D). Les K_D ont été calculés à partir des équations présentées en 2.2.2, en assumant que les systèmes étudiés ici sont des hétérodimères (ce qui est vrai pour l'hétérodimère c-Myc*/Max*, mais toujours inconnu pour c-Myc*/Mid2s- β) (Naud JF, et al. 2003). Ces expérimentations montrent que le Mid2s- β possède des déterminants structuraux permettant de former un complexe de stabilité thermodynamique plus élevée que celle du complexe c-Myc*/Max*.

A



B



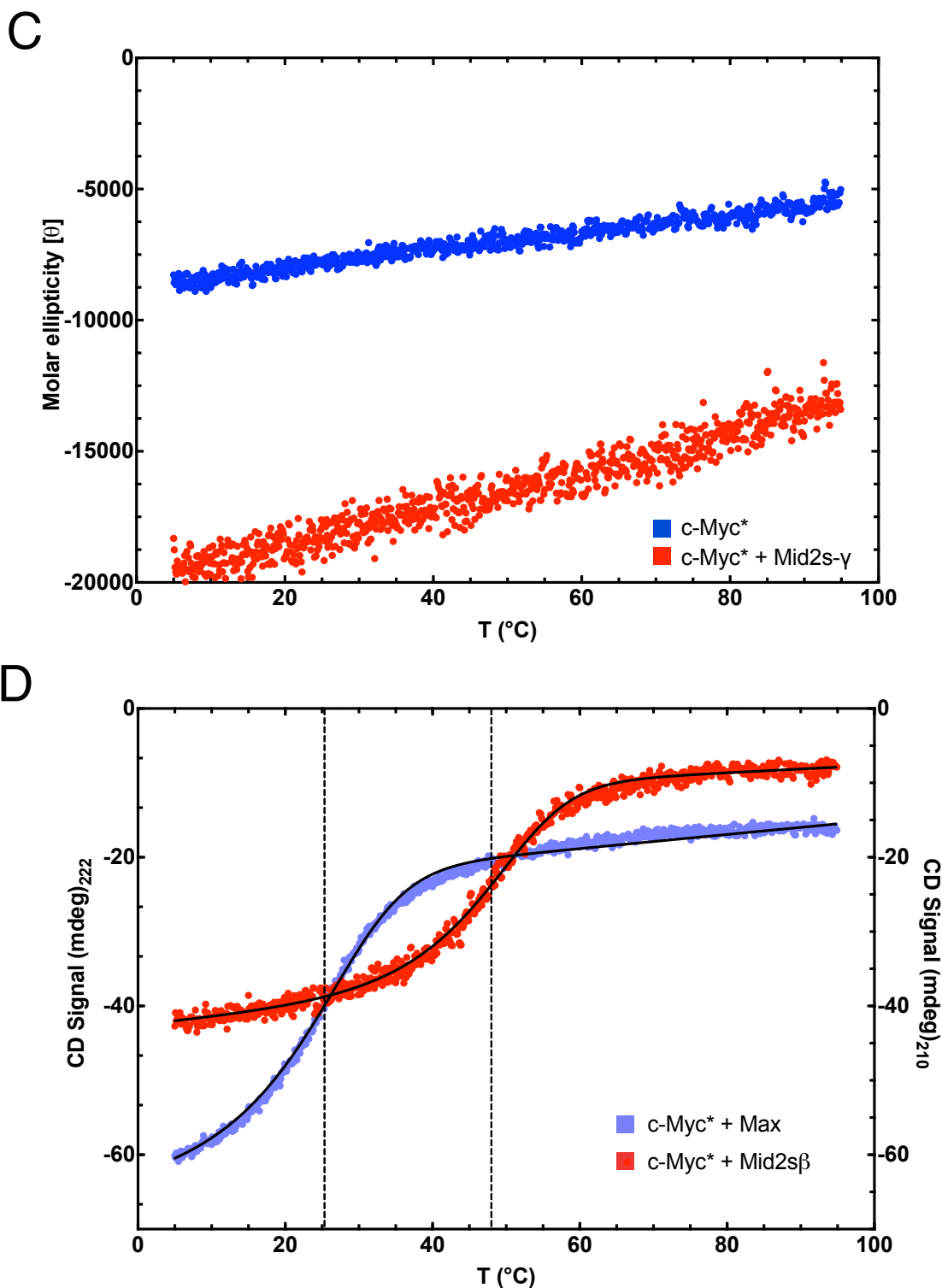


Fig.16 Caractérisation des complexes c-Myc/Mid2s/Max par dichroïsme circulaire

(A) Dénaturations thermiques de c-Myc* et c-Myc*/Mid2s- α à 209nm (B) Dénaturations thermiques de c-Myc* et c-Myc*/Mid2s- β à 210nm (C) Dénaturations thermiques de c-Myc* et c-Myc*/Mid2s- γ à 210nm (D) Dénaturations thermiques des complexes c-Myc*/Max* et c-Myc*/Mid2s- β à 222nm et 210nm respectivement. Le T_m du complexe c-Myc*/Max* est à 25.3°C, le T_m du complexe c-Myc*/Mid2s- β est à 48°C

3.1.3 Mid2s- β perturbe le complexe c-Myc/Max : ADN pour favoriser la formation du complexe Max/Max : ADN

En observant les résultats présentés dans les précédentes sections, il est possible de proposer que le peptide Mid2s- β pourrait séquestrer c-Myc* de son complexe avec Max*, en absence d'ADN. Comme présenté dans l'introduction, il importe de pouvoir empêcher la liaison de c-Myc sur l'ADN, pour pouvoir exploiter le talon d'Achille des cellules cancéreuses qu'est l'addiction oncogénique. Pour ce faire, des expérimentations EMSA (Electrophoretic Mobility Shift Assays) ont été effectuées afin d'observer l'effet du peptide Mid2s- β sur la liaison à l'ADN de c-Myc. Plus précisément, une titration du peptide Mid2s- β a été effectuée sur le complexe c-Myc*/Max*/ADN. Dans ce cas-ci, l'ADN utilisé était soit un oligomère fluorescent (iFluor) palindromique de séquence E-box canonique (CACGTG) ou non-spécifique (E-box brouillé – GGATCC). De plus, la même titration a été effectuée avec les peptides Mid2s- α et Mid2s- γ (Fig. 18A et C respectivement). Pour la titration avec le peptide Mid2s- β (Fig. 18B), une diminution du complexe hétérodimérique c-Myc*/Max*/ADN est apparente au profit de l'apparition du complexe homodimérique Max*/Max*/ADN autant avec la séquence E-box canonique qu'avec la séquence non-spécifique. Il est à noter qu'aucun complexe de plus haut poids moléculaire pouvant correspondre à c-Myc*/Max*/Mid2s- β /ADN ou de plus faible poids moléculaire pouvant correspondre à c-Myc*/Mid2s- β /ADN ne sont apparents.

Ces résultats sont supportés par le fait que par expérimentations CD, aucun complexe entre l'ADN et c-Myc*/Mid2s- β ne peut être formé (Fig.17A). Cette figure montre que le complexe c-Myc*/c-Myc*/ADN est perturbé au profit du complexe c-Myc*/Mid2s- β , en se fiant aux T_{MS} apparents. L'existence du complexe c-Myc*/c-Myc*/ADN a été reporté à plusieurs reprises, notamment par EMSAs et par dichroïsme circulaire (Fig.17B) (Jung LA, et al. 2016) (Maltais L, et al. 2017). Sur cette figure, il est possible de voir l'apparition d'une courbe de dénaturation coopérative correspondant au complexe c-Myc*/c-Myc*/Ebox, ce qui concorde avec les résultats obtenus dans nos expérimentations (Fig.17A). Les EMSAs montrant une bande correspondant au complexe c-Myc*/c-Myc*/Ebox ont été effectués dans des conditions différentes que celles utilisées dans nos expérimentations, notamment la température et le tampon de migration utilisé (Jung LA, et al. 2016).

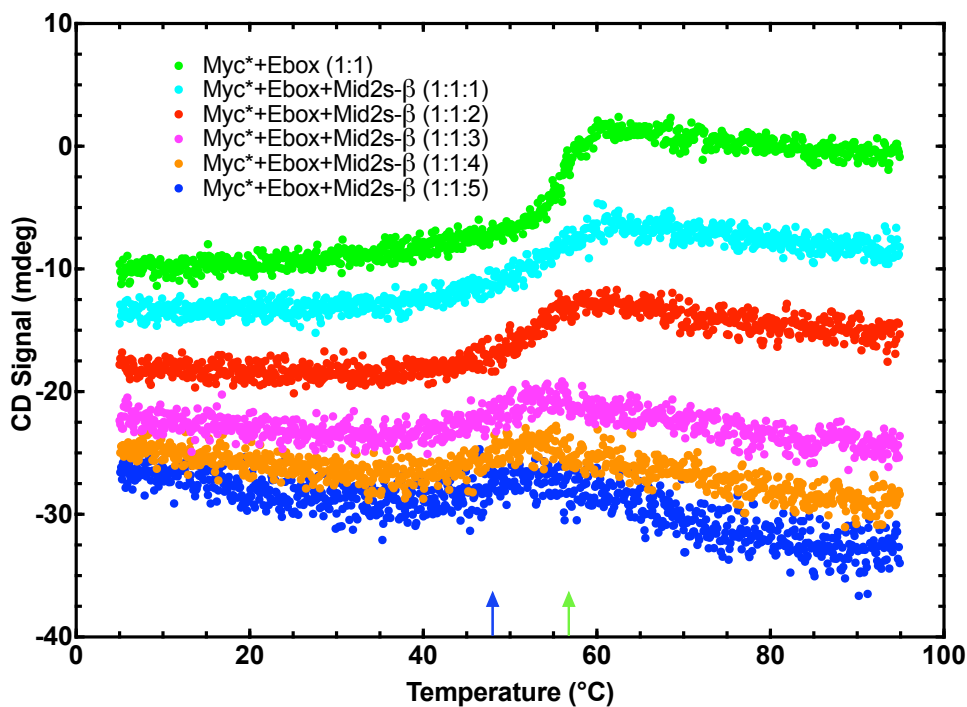
L'évolution du système c-Myc*/c-Myc*/Ebox vers c-Myc*/Mid2s- β par CD est illustré dans la Fig.17A. Il est possible d'observer ce phénomène en suivant l'évolution du T_M en fonction des incréments de peptide Mid2s- β ajouté au complexe c-Myc*/c-Myc*/Ebox. La flèche verte sur la figure 17A indique le T_M apparent de ce complexe (56°C), qui correspond aussi au T_M apparent de la figure 17B, tirée du manuscrit de Maltais L, et al. 2017. Il est possible de noter qu'après l'ajout d'un excès de 5 fois la quantité en équimolaire de peptide Mid2s- β , le T_M apparent est à 48°C (flèche bleue), ce qui correspond au T_M normalement obtenu pour le complexe c-Myc*/Mid2s- β (voir 3.1.2). Le modèle associé aux résultats présentés ici est exposé dans la figure 17D.

L'évolution du système c-Myc*/Max*/Ebox vers d'autres populations de complexes (comprendre c-Myc*/Mid2s- β , Max*/Max*/Ebox et c-Myc*/Max*/Ebox) par CD est illustré dans la Fig.17C. Il était possible de suivre l'évolution de c-Myc*/c-Myc*/Ebox vers c-Myc*/Mid2s- β dans la Fig.17A à l'aide du T_M puisque le système passait d'un dimère à un autre, sans avoir de compétition (l'homodimère c-Myc*/c-Myc* Apo ne pouvant exister en quantité suffisamment grande à cause de son K_D très faible; Maltais L et al. 2017). Or, dans l'expérience présentée en 17C, beaucoup de populations sont présentes, et le T_M en dichroïsme circulaire représentera une moyenne des T_M des complexes en solution. Il est donc difficile d'affirmer avec grande précision l'équilibre des populations qui est présenté dans la figure 17C. Cependant, le T_M moyen augmente en fonction des incréments de Mid2s- β ajoutés, suggérant que le peptide Mid2s- β stabilise une population dans l'ensemble présent dans la cuvette CD. La même expérimentation a été répétée en 3.1.4 et en 3.1.5 par RMN afin d'avoir plus de détails à cet égard.

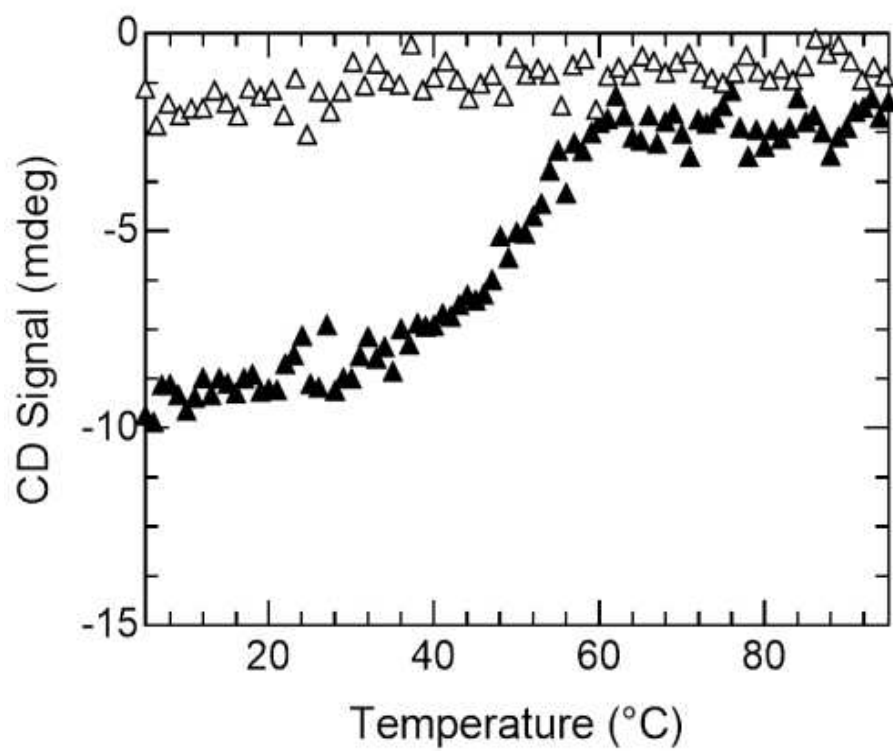
Il est possible d'observer, en EMSAs, que l'évolution du complexe c-Myc*/Max*/ADN vers la forme Max*/Max*/ADN suite à la titration avec le peptide Mid2s- β n'est pas complète. Cela peut être expliqué par le fait que l'hétérodimère c-Myc*/Max* et l'homodimère Max*/Max* possèdent la même affinité pour l'ADN (Guo, J et al. 2014) (Maltais L, et al. 2017). Une faible évolution du complexe c-Myc*/Max*/ADN vers la forme Max*/Max*/ADN suite à la titration avec le peptide Mid2s- α est présente. Il est à noter que Mid2s- α possède quelques résidus en commun avec Mid2s- β , expliquant la faible interaction observée (Fig. 18A). Ceci concorde avec

la faible formation de complexe entre Mid2s- α et c-Myc* en dichroïsme circulaire (Fig. 16A). De plus, il n'y a pas d'association notable entre Mid2s- γ et c-Myc*, puisque que le complexe c-Myc*/Max*/ADN n'est pas perturbé (Fig. 18C). Afin de valider que la perturbation du complexe c-Myc*/Max*/ADN et la formation concomitante du complexe Max*/Max*/ADN (Fig. 17E) est la résultante de la formation d'un complexe spécifique c-Myc*/Mid2s- β , les titrations de peptides par EMSA ont été répétées avec le complexe c-Myc*V394D/Max*/ADN (Fig. 18D). Comme présenté en 1.3.2, le mutant c-Myc*V394D (mutation dans le HLH de c-Myc*) abolit l'interaction entre c-Myc et Miz-1 (Peukert K, et al. 1997) (Bédard M, et al. 2016). En effectuant l'expérimentation EMSA avec ce mutant (Fig.18D), il est possible d'observer que le complexe c-Myc*V394D/Max*/ADN se forme, et le peptide Mid2s- β n'a plus la capacité de perturber le complexe comme auparavant (Fig.18B). Ce résultat était prévisible dans la mesure où une interaction Miz-1/c-MycV394D n'est pas censée être observable, et vient valider que les déterminants structuraux de Miz-1 responsables de son interaction avec le b-HLH-LZ de c-Myc sont situés dans la région englobant les résidus 660 à 692.

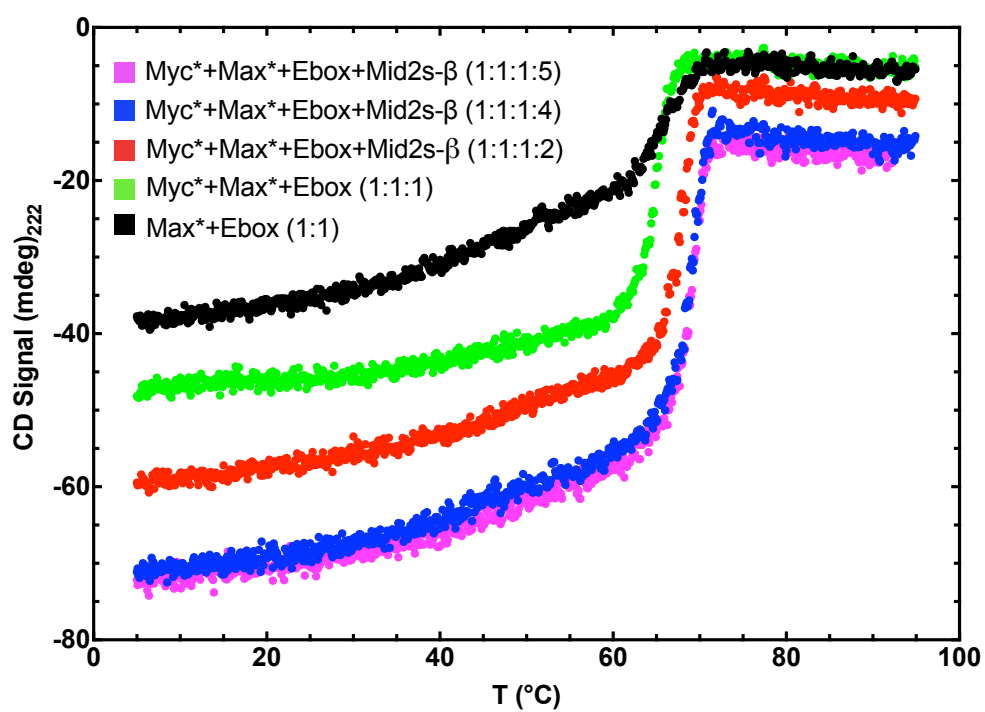
A



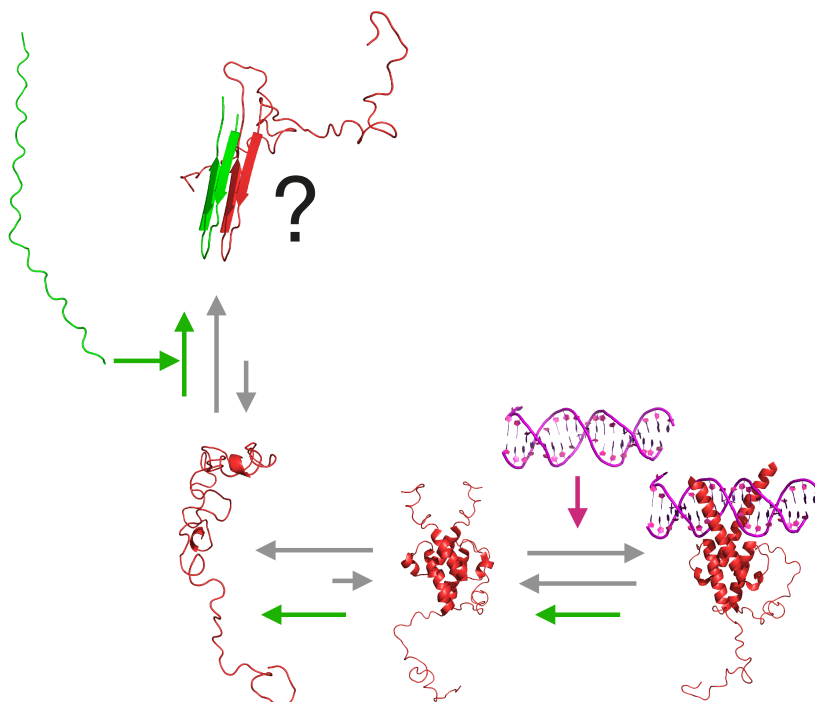
B



C



D



E

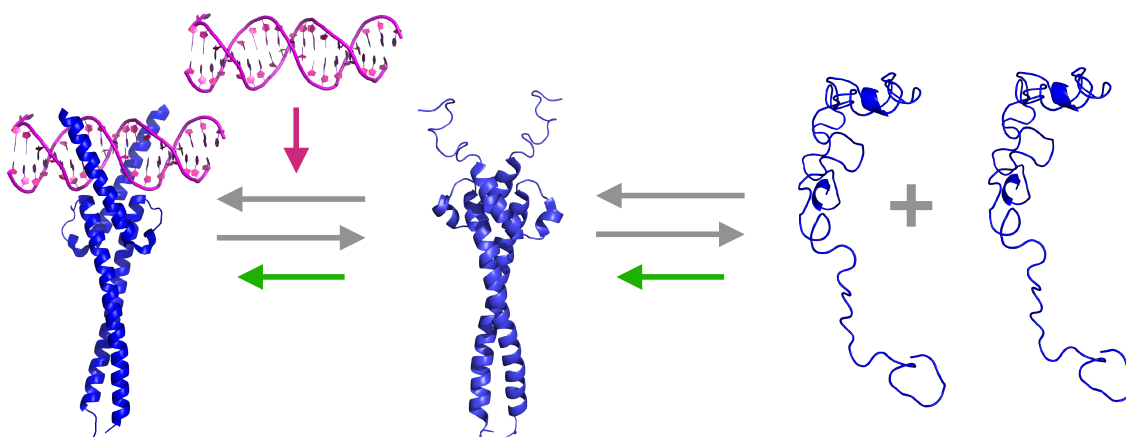


Fig.17 Titration du complexe c-Myc*/ADN avec Mid2s-β et modèle associé

(A) Dénaturations thermiques à 222nm du complexe c-Myc*/ADN à (20μM ADNdb, 40μM c-Myc*), titré avec le peptide Mid2s-β à des incréments de 10μM jusqu'à une concentration de 50μM. La flèche verte indique le T_m du complexe c-Myc*/c-Myc*/Ebox, la flèche bleue indique le T_m du complexe c-Myc*/Mid2s-β. (B) Dénaturation thermique de c-Myc*/ADNnon-spécifique (triangles vides) et de c-Myc*/c-Myc*/Ebox (triangles noirs). Figure tirée de Maltais L, et al. 2017 avec l'approbation de l'auteur. (C) Modèle associé à l'équilibre entre le complexe c-Myc*/c-Myc*/ADN (rouge) et c-Myc*/Mid2s-β (vert). (D) Modèle associé à l'homodimérisation de Max* (bleu) pour la liaison à l'ADN.

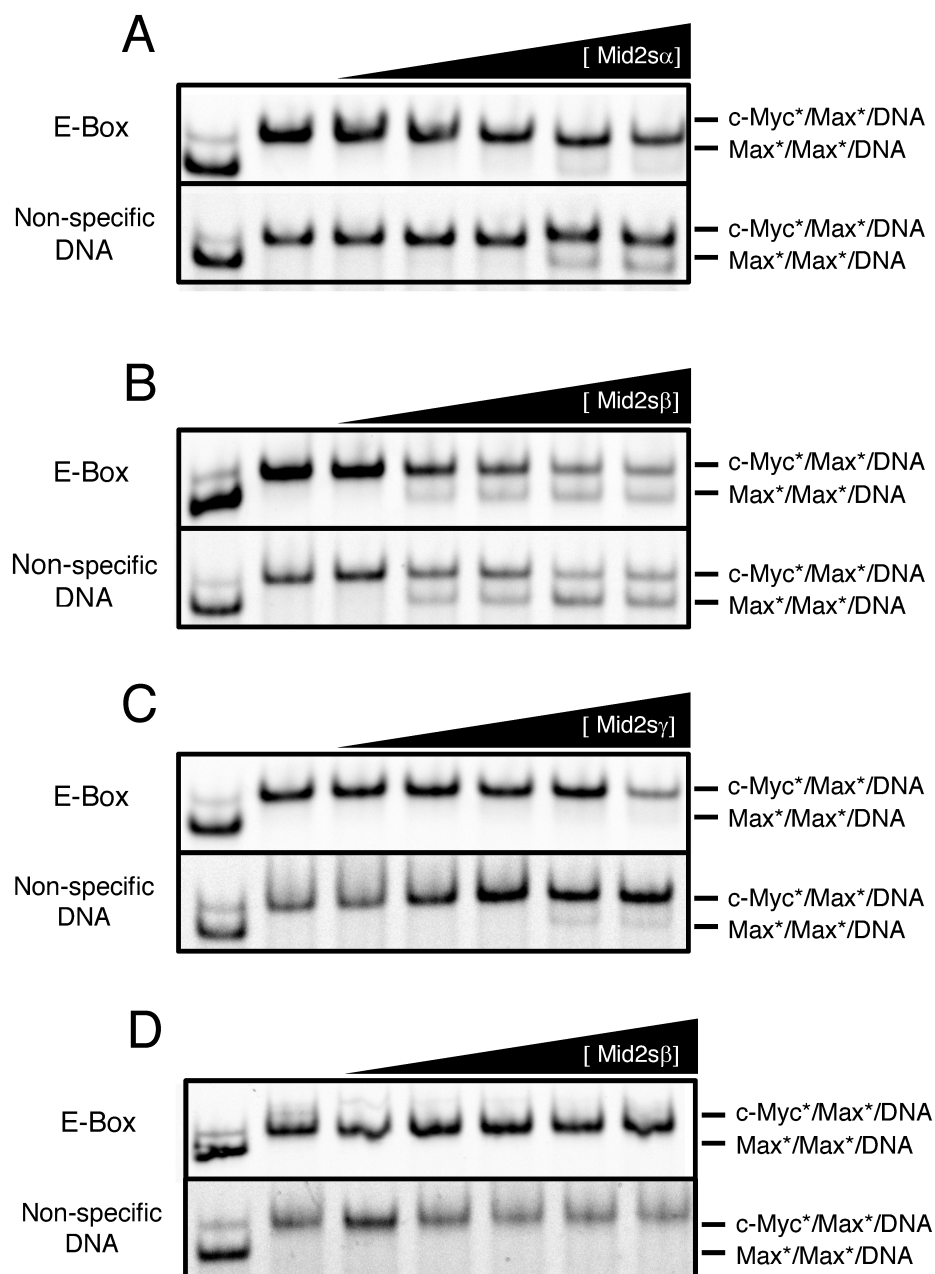


Fig.18 Caractérisation des complexes c-Myc/Mid2s/Max par EMSA

(A-B-C) Expérimentations EMSA suivant la titration du complexe c-Myc*/Max*/ADN avec les peptides Mid2s- $\alpha/\beta/\gamma$. Les deux premiers puits à gauche contiennent le complexe Max*/Max*/ADN (1000nM Max*) et le complexe c-Myc*/Max*/ADN (500nM Max*, 500nM c-Myc*). Les puits suivants contiennent le complexe c-Myc*/Max*/ADN titré avec 500nM; 5000nM; 10 000nM; 50 000nM; 100 000nM Mid2s- $\alpha/\beta/\gamma$. La concentration d'ADN marqué dans chaque puit est de 250nm (double-brin). **(D)** Titration du complexe c-Myc*V394D/Max*/ADN avec Mid2s- β comme expliqué précédemment.

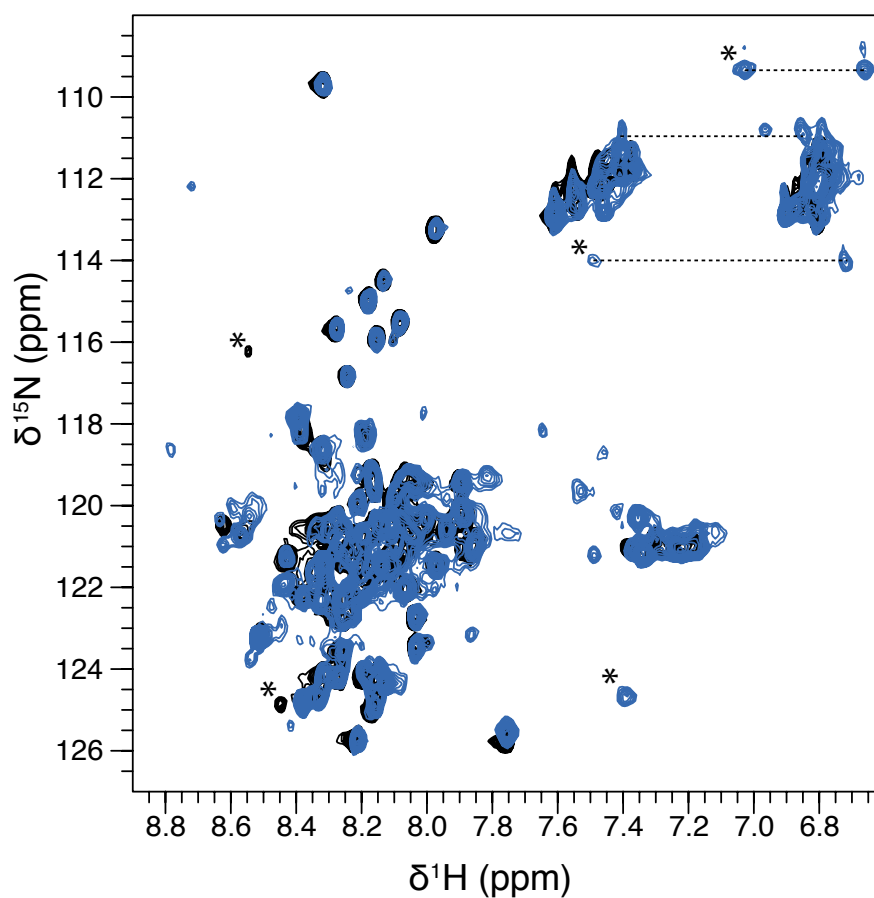
3.1.4 Mid2s- β séquestre c-Myc pour favoriser l'homodimère de Max

Ayant déterminé que la région Mid2s- β de Miz-1, englobant les résidus 660 à 692, contient les déterminants structuraux afin de former un complexe stable avec c-Myc*, il semblait important de comprendre la nature compétitive et spécifique du dit complexe. Il était possible d'avoir ce type d'information à l'aide de titrations par ^1H - ^{15}N -HSQC (voir section 2.4.4). D'abord, en utilisant le bHLH-LZ de Max* marqué ^{15}N , il était possible d'observer l'hétérodimérisation avec c-Myc* (non-marqué) via l'apparition de nouveaux pics caractéristiques dans le ^1H - ^{15}N -HSQC. Il est à noter que Max* possède une affinité plus grande pour l'homodimérisation avec lui-même que pour l'hétérodimérisation avec c-Myc* (Maltais L, et al. 2017) (Jung LA, et al. 2016). Il était donc nécessaire d'utiliser un léger excès de c-Myc* (3 fois plus - 150 μM) afin de voir apparaître la forme hétérodimérique dans le ^1H - ^{15}N -HSQC de Max*. Donc, dans la figure 19A, on peut voir la forme homodimérique de Max* (en noir) et la forme hétérodimérique c-Myc*/Max* (en bleu). Comme on pourrait s'y attendre, l'hétérodimérisation amène des changements dans les déplacements chimiques observés. Comme observé dans les expériences de dichroïsme circulaire, il y a présence d'une population de Max* homodimérique et d'hétérodimère Max*/c-Myc*, et cela se traduit dans le ^1H - ^{15}N -HSQC (deux populations de pics dans le spectre). Cependant, il est quand même possible d'annoter les nouveaux pics apparus, puisqu'ils appartiennent à l'hétérodimère c-Myc*/Max*, et donc font office de pics caractéristiques de l'hétérodimérisation. L'apparition de chaînes latérales caractéristiques de l'hétérodimère, ainsi que certains nouveaux pics (annotés par des astérisques) (Fig. 19A) nous permettent de suivre l'évolution de cette forme hétérodimérique et constituent la preuve de la présence de l'hétérodimère. Ces chaînes latérales ont été assignées précédemment. Les deux chaînes latérales annotées correspondent aux résidus Q42 de Max* (pour la chaîne latérale à 109.5ppm et Q63 pour la chaîne latérale à 113.5ppm) (Sauvé S, et al. 2004). En titrant avec du peptide Mid2s- β à 250 μM et 500 μM (5 et 10 équivalents), il est possible d'observer une disparition des pics caractéristiques de l'hétérodimère c-Myc*/Max* au profit d'une apparition de pics caractéristiques de l'homodimère Max*/Max* (Fig. 19B).

Cela nous montre donc que le peptide Mid2s- β est capable de séquestrer c-Myc* de l'hétérodimère c-Myc*/Max* afin de favoriser la forme homodimérique Max*/Max*, ce qui vient valider les résultats observés dans les expérimentations EMSA (voir 3.3). Étant donné que ces

expériences RMN ne nous donnent que des informations sur l'état du bHLH-LZ de Max* dans ces conditions, il n'est pas possible d'affirmer ce qu'il advient du complexe c-Myc*/Mid2s- β . Il est cependant possible de valider ce qui était observé en dichroïsme circulaire et en essais de retards sur gel – il n'y a pas de formation apparente d'un complexe Max*/Mid2s- β , appuyant donc le fait que Mid2s- β reconnaît spécifiquement c-Myc*.

A



B

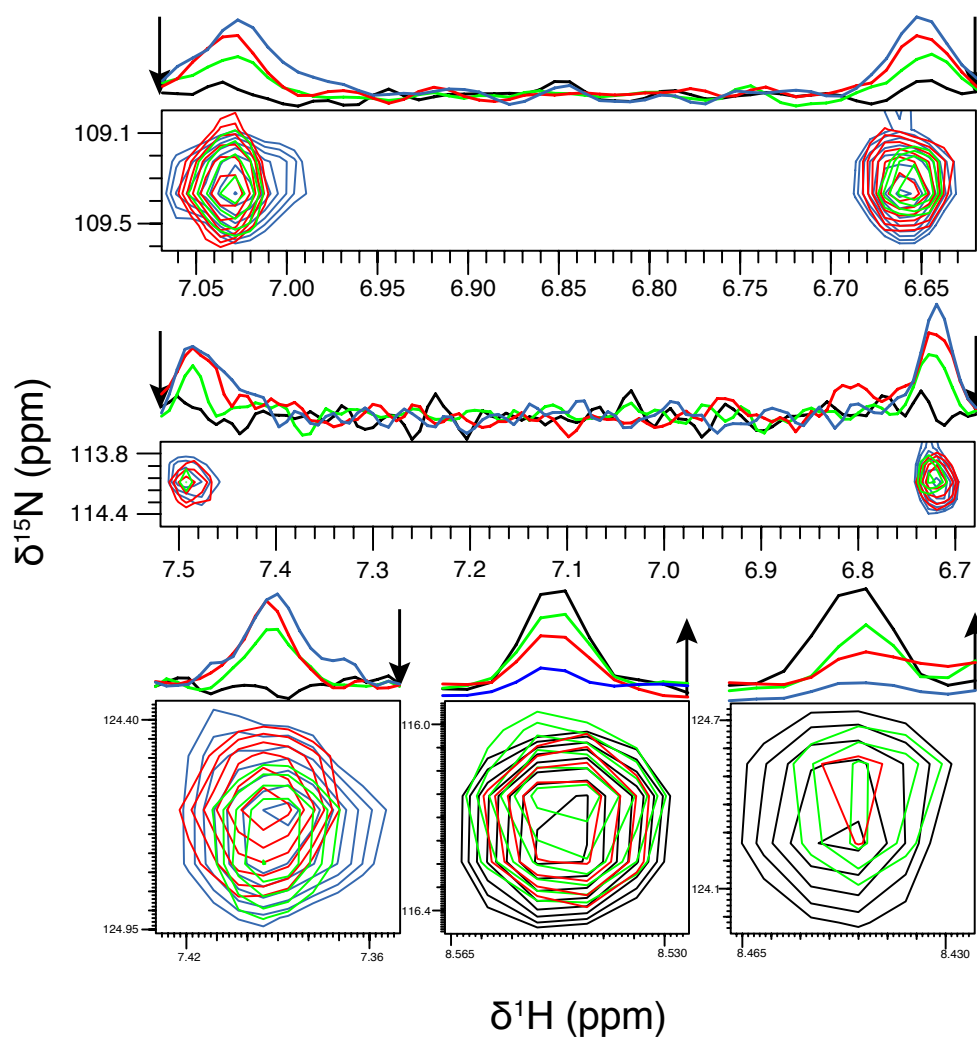


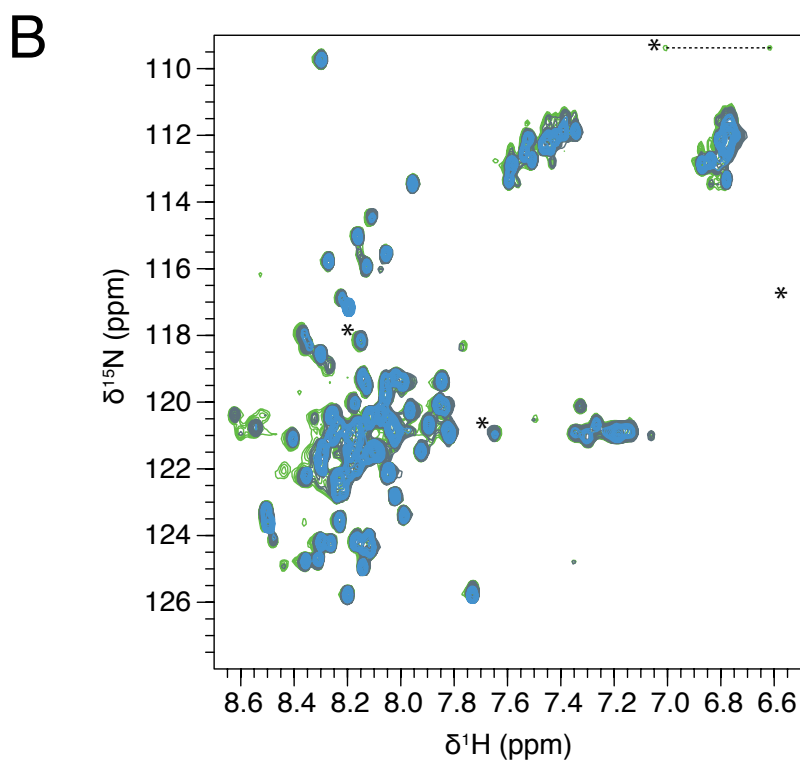
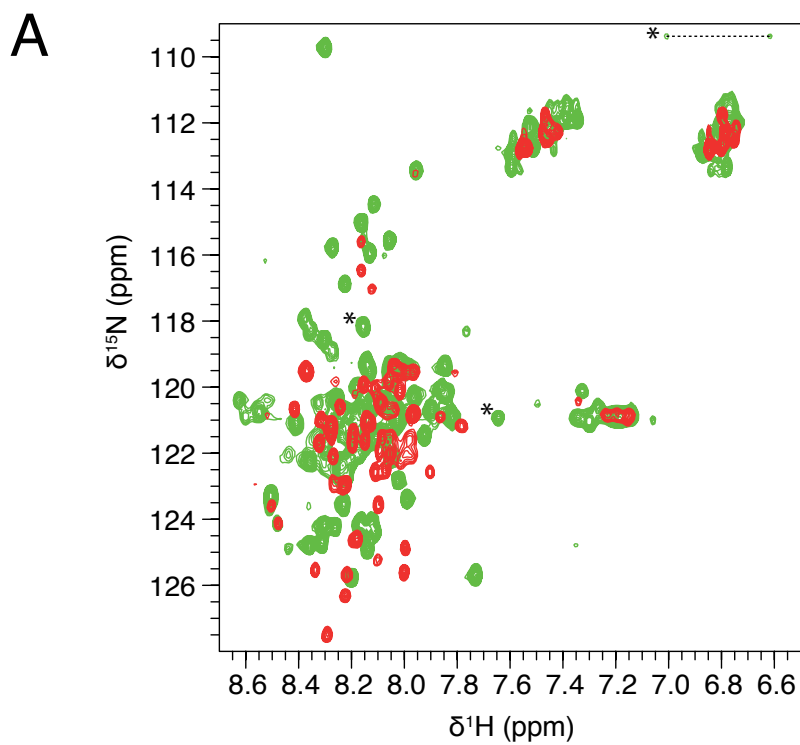
Fig.19 Titration par ^1H - ^{15}N -HSQC du complexe c-Myc*/ ^{15}N -Max* avec Mid2s- β
(A) ^1H - ^{15}N -HSQC de Max* homodimérique (Noir), ^1H - ^{15}N -HSQC de l'hétérodimère c-Myc*/Max (Bleu) à un excès de 3 équivalents de c-Myc*. Les pics caractéristiques de l'hétérodimère sont annotés d'un astérisque. **(B)** Focus sur les pics annotés, ainsi que leur intensité relative.

3.1.5 Mid2s- β compétitionne avec Max pour former un complexe exclusif avec c-Myc

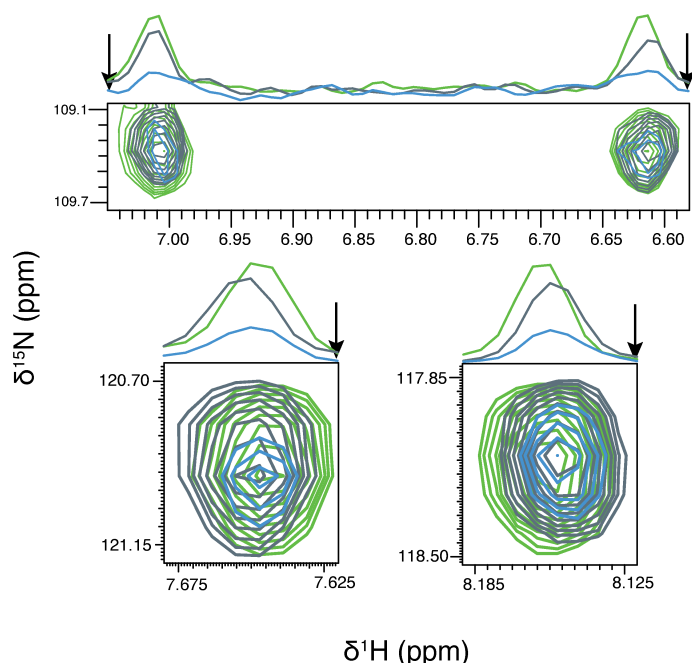
Afin de comprendre ce qu'il advient du complexe c-Myc*/Mid2s- β , il convenait donc d'effectuer les expériences RMN réciproques de celles effectuées en 3.1.4. En d'autres mots, il incombait de faire la titration, mais cette fois ci en observant le bHLH-LZ de c-Myc marqué ^{15}N . Dans la figure 20A, un ^1H - ^{15}N -HSQC de c-Myc* sous sa forme libre (apo, en rouge) a été effectué. Contrairement à Max*, c-Myc* n'a qu'une très faible affinité pour homodimériser. Il est donc possible de voir une faible dispersion des pics du HSQC, diagnostic d'un repliement faible qui concorde avec la nature intrinsèquement désordonnée de c-Myc (Kwan AH, et al. 2011) (voir 1.4.3). Un excès de Max* n'était donc pas nécessaire afin d'obtenir la forme hétérodimérique c-Myc*/Max*. En ajoutant 1 équivalent de Max*, on obtient le ^1H - ^{15}N -HSQC en vert dans la figure 20A, qui correspond à une population uniforme de complexe c-Myc*/Max*. Certains pics diagnostiques de l'hétérodimère c-Myc*/Max* ont été sélectionnés (astérisques) afin de suivre leur évolution durant la titration. En ajoutant des incréments de Mid2s- β , ici 5 et 10 équivalents, on peut voir l'intensité des pics diagnostiques sélectionnés diminuer fortement (Fig. 20B-C). L'intensité des pics diminuée ainsi pourrait apparaître comme étant une conséquence d'échange lent (voir 2.4.4). Cependant, les pics concernés ne réapparaissent pas en fonction des incréments dans un spectre de population c-Myc*/Mid2s- β . En temps normal, la formation d'un complexe c-Myc*/Mid2s- β entrainerait l'apparition d'une nouvelle population de déplacements chimiques dans le ^1H - ^{15}N -HSQC. Cependant, la titration avec le peptide Mid2s- β n'entraîne que la perte de signal de l'hétérodimère. Cela suggère que Mid2s- β , en interagissant, séquestre c-Myc* spécifiquement en l'entraînant dans une forme de haut poids moléculaire, puisqu'une telle perte de signal est dû à un *tumbling* lent. Cela n'est pas à confondre avec de l'échange lent, puisque la perte de signal serait associée à un gain de signal dans une autre population dans le ^1H - ^{15}N -HSQC, et ce n'est pas ce qui est observé ici. La diffusion rotationnelle *tumbling* lente peut être attribuée à des espèces de haut poids moléculaire pouvant être formées facilement par les IDPs, notamment des structures *cross*- β de type amyloïde (voir 1.4.4). En plus d'être de plus haut poids moléculaire, ces structures forment souvent des hydrogels en solution, réduisant drastiquement le *tumbling* des molécules – expliquant alors la perte de signal engendrée dans les résultats obtenus par ^1H - ^{15}N -HSQC.

Étant donné la nature intrinsèquement désordonnée de c-Myc et du peptide Mids- β (structure PPII associée aux IDPs, voir 1.4.3), il incombait de vérifier de manière computationnelle si le bHLH-LZ de c-Myc et le peptide Mids- β peuvent former des structures *cross- β* de type amyloïde, venant alors valider les résultats obtenus par ^1H - ^{15}N -HSQC. AGGRESCAN est un outil bioinformatique qui permet d'anticiper la formation d'agrégats par des polypeptides, notamment sous forme de structures *cross- β* de type amyloïde. Ce logiciel permet de prédire des segments ayant une propension pour l'agrégation dans une structure primaire, l'analyse de l'effet de mutations sur la propension à l'agrégation et la comparaison des propriétés d'agrégation entre différentes protéines. L'algorithme se base sur une échelle de propension à l'agrégation des acides aminés naturels tiré de résultats expérimentaux *in vivo* et sur l'hypothèse que certains courts segments de séquence spécifique modulent la propension à l'agrégation (Conchillo-Solé O, et al. 2007). Les résultantes des analyses des structures primaires de c-Myc et de Miz-1 sont présentées à la Fig.19D et 19E, respectivement. Souligné en bleu, dans la Fig.20D, il est possible d'observer la structure primaire du bHLH-LZ de c-Myc, avec un segment en rouge ayant une propension à former des agrégats de structure *cross- β* de type amyloïde, prédit à l'aide d'AGGRESCAN. Ce segment correspond à l'hélice 2 du bHLH-LZ de c-Myc. Il est à noter que ce segment contient le résidu V394 de c-Myc discuté à maintes reprises dans ce manuscrit. Souligné en vert, dans la Fig.20E, il est possible d'observer la structure primaire du peptide Mid2s- β tiré de Miz-1, avec un segment en rouge ayant une propension à former des agrégats de structure *cross- β* de type amyloïde, prédit à l'aide d'AGGRESCAN. Ce segment correspond à la quasi-totalité du peptide Mid2s- β . Ces résultats computationnels, combinés aux résultats obtenus notamment par ^1H - ^{15}N -HSQC et dans les précédentes sections, nous permettent d'émettre un modèle théorique général de séquestration de c-Myc* par Mid2s- β sous forme de feuillets amyloïdes (ou autre) pour l'empêcher de lier l'ADN sous forme d'hétérodimère avec Max*, présenté à la Fig.21. Dans ce modèle, on observe l'équilibre entre les états Max* homodimérique et c-Myc* homodimérique liés ou non à l'ADN qui peuvent hétérodimériser pour lier l'ADN. La séquestration de c-Myc* par Mid2s- β favorise la formation de Max* homodimérique, et ce qui au final permettrait d'inhiber l'hétérodimérisation, la liaison de c-Myc à l'ADN et en dernier lieu la tumorigénèse des cellules

cancéreuses. Ce modèle général projette alors les bases d'une nouvelle stratégie pour la génération d'inhibiteurs du bHLH-LZ de c-Myc, jusqu'à maintenant insoupçonnée.



C



D

MPLNVSFTNRNYDL DYDSVQPYFYCDEEENFYQQQQQSELQPPAPSEDIW**KKFELL**PTPPLS
 PSRRSGLCSP**SYVAVT**PFSLRGDNDGGGGSFSTADQLE**MVTELL**GDMVNQSFICDPDDETFI
KNIIQDCMWSGFSAAAKLVSEKLASYQAARKDSGSPNPARGHSVC**STSSLYLQDL**SAAASEC
 IDPSVVFYPLNDSSSPKSCASQDSSAFSPSSDILLSSTESSPQGSPEPLVLHEETPPTTSSDS
 EEEQEDEEEIDVVSVEKRQAPGKRSESGSPSAGGHSKPPHSPLVLKRCHVSTHQHNYAAPP
 TRKDYPAAKRVKL**DSVRV**LQRISNNRKCTSPRSSDTEENVKRRTHNVLERQRRNELKRSFFAL
 RDQIPELENNEKAP**KVILK**KATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLRNSCA

E

MDFPQHSQHVLEQLNQQRQLGL**LC**DCTFVDGVHFKAHKAVLAACSEYFKMLFVDQKD VVH
 LDISNAAGL**GQVLEFMYTAKL**SLSPENVDDV**LAVATFLQM**QDIITACHALKSLAEPATSPGGNA
 EALATEGGDKRAKEEKVATSTLSRLEQAGRSTPIGPSRDLKEERGGQAQSAASGAETEAD
 APREPPPVELKPDPTSGMAAAEAEALSESSESEQEMEVEPARKGEEEEEQKEQEEQEEGAGPAE
 VKEEGSQLENGEAPEENENEESAGTDSGQELGSEARGLRSGTYGDRTESKAYGSGVIHKCEDC
 GKEFTHTGNFKRHIRIHTGEKPFSCRECSKAFSDPAACKAHEKTHSPLKPYGCEECGKS**YRLI**
SLLNLHKKRHSGEARYRCEDCGKLFTTSGNLKRHLVHSGEKPYQCDYCGRSFSDPTSKMR
 HLETHDTDKEHKCPHCDKKFNQVGNL**KAHLKI**HIADGPLKCRECGKQFTTSGNLKRHLRIHS
 GEKPYVCIHCQRQFADPGALQRHVRIHTGEKPCQC**VMCGKAFTQASS**LIAHVRQHTGEKPYV
 CERCGKRFVQSSQLANHIRHHDNIRPHKCS**VCSKAFVNVGDLSKHIII**HTGEKPYLCDKCGRG
 FNRVDNLRSHVKT VHQGKAGIKILEPEEGSE**VSVVTVD**DMVTLATEALAA**TA**VTQLTVVPVGAA
VTADETEVLKAEISKAVKQVQEEDPNTH**LYACD**SCGDKFLDANSLAQHVRIH**TAQALVM**FQTD
 ADFYQQYGPGGTWPAGQVL**QAGEL**VFRPRDGAEGQPALAETSPTAPECPPAE

Fig.20 Titration par ^1H - ^{15}N -HSQC du complexe ^{15}N c-Myc*/Max* avec Mid2s- β

(A) ^1H - ^{15}N -HSQC de c-Myc* dans la forme libre (Rouge), ^1H - ^{15}N -HSQC de l'hétérodimère c-Myc*/Max (Vert). Les pics caractéristiques de l'hétérodimère sélectionnés sont annotés d'un astérisque. (B) ^1H - ^{15}N -HSQC associés aux points de la titration avec Mid2s- β (5 éq : Gris, 10 éq : Bleu). (C) Focus sur les pics annotés, ainsi que leur intensité relative. (D) Propension à l'agrégation de type amyloïde (en rouge) de la structure primaire de c-Myc prédit par Aggrescan. En souligné, c-Myc* et encadré V394. (E) Propension à l'agrégation de type amyloïde (en rouge) de la structure primaire de Miz-1 prédit par Aggrescan. En

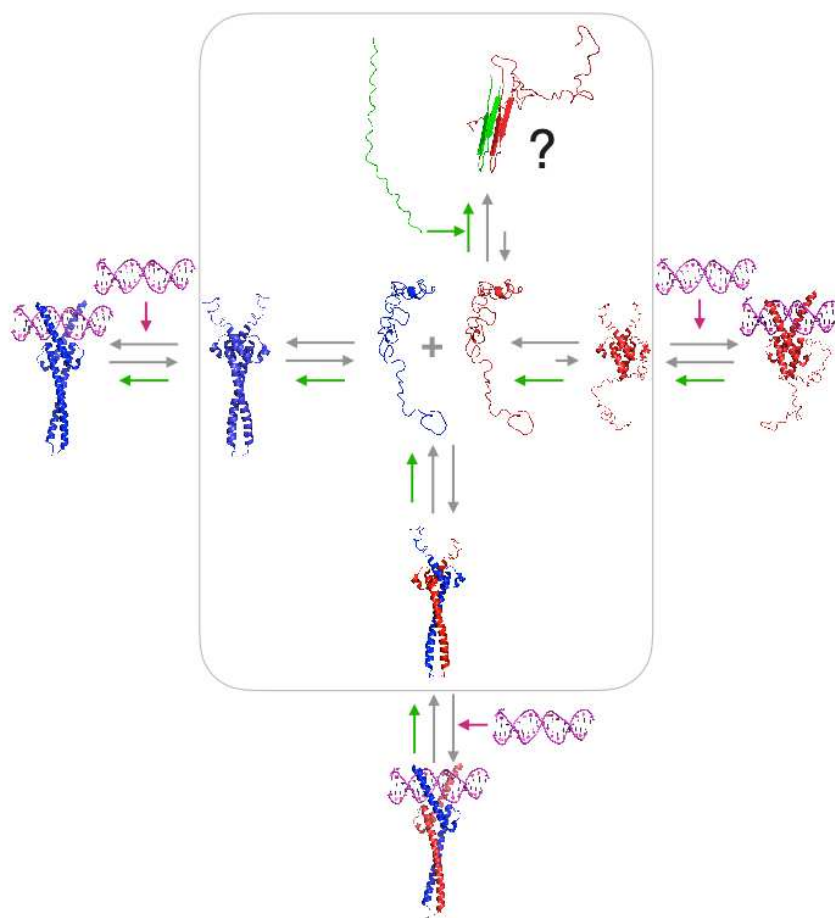


Fig.21 Modèle général de séquestration du bHLH-LZ de c-Myc par Mid2s-β

Modèle général de séquestration du bHLH-LZ de c-Myc par Mid2s-β, regroupant les modèles présentés à la figure 15C et 15D. Le modèle présente les mécanismes validés par les expérimentations effectuées ici, ainsi que le mécanisme d'agrégation de type amyloïde suggéré par les méthodes computationnelles (Aggrescan). Les protéines en vert correspondent à Mid2s-β, les bHLH-LZ en bleu correspondent à Max et les bHLH-LZ en rouge correspondent à c-Myc.

4. Discussion

4.1 Contexte et portée de l'étude

L'activation de l'activité oncogénique de c-Myc passe d'abord et avant tout par des changements génétiques, telles que des mutations dans des voies de signalisations importantes ou des translocations. Comme exposé précédemment, il se trouve que c-Myc est un des oncogènes les plus fréquents dans l'apparition de tumeurs. Suite à des preuves de concept en modèles murins, il a été démontré que c-Myc serait une excellente cible pour induire l'apoptose et la différenciation dans les cellules en état d'addiction oncogénique, et donc amener la régression de tumeurs c-Myc dépendantes. Plusieurs inhibiteurs prototypiques sous forme de petites molécules ont été développés à l'aide de criblage à haut débit dans les dernières années, sans succès apparent. Il convenait donc d'essayer une nouvelle approche.

Le fait que c-Myc soit une protéine intrinsèquement désordonnée fait en sorte que l'utilisation de design rationnel d'inhibiteurs, tel que le criblage à haut débit, ne peut fonctionner à proprement dit puisque la cible ne possède pas de site de liaison explicite. L'utilisation de mécanismes d'inhibition de c-Myc présents dans la nature, tel que celui observé avec Miz-1 lorsqu'on empêche sa dégradation avec un inhibiteur de HUWE1 (Peter S, et al. 2014), était un excellent point de départ dans le développement d'inhibiteurs prototypiques de c-Myc.

Ce mémoire visait donc une étude structurale et biophysique du complexe c-Myc/Mid2 dans le but d'élucider sa capacité en tant qu'inhibiteur prototypique. Les résultats obtenus dans les précédentes sections nous offrent des informations fortement pertinentes sur un nouveau mécanisme d'inhibition de c-Myc. La séquestration spécifique de c-Myc via une région minimale du domaine Mid2 de Miz-1 et l'élucidation de ses déterminants moléculaires constituent donc une avancée considérable dans la validation d'un mécanisme de répression de la tumorigénèse via le c-Myc/Mid2s- β .

4.2 Discussion sur la séquestration de c-Myc par les peptides Mid2s

Dans les résultats présentés au chapitre 3, il a été démontré que la section englobant les résidus 660-692 du facteur de transcription à doigts de zinc Miz-1, dans le domaine Mid2 forme un complexe spécifique avec c-Myc. Cette section, d'une longueur de 32 acides aminés, correspond au peptide surnommé Mid2s- β . Lorsque séquestré dans ce complexe, le b-HLH-LZ de c-Myc adopte une structure différente de celle adoptée en hétérodimère avec Max, et rend la protéine incapable de lier l'ADN. Il est à noter que cette structure est de haut poids moléculaire et semble former des structures *cross- β* de type amyloïde (voir 3.1.5). D'abord, ces conclusions diffèrent avec les résultats obtenus par le groupe de recherche de Peukert en 1997 dans le sens où la région prédite d'interaction entre Miz-1 et c-Myc*/Max* correspondait au peptide Mid2s- γ (selon notre notation) et s'effectuait via des structures hélicales. Cependant, nos résultats montrent que cette interaction se fait via le peptide Mid2s- β et via des structures *cross- β* de type amyloïde.

Le groupe de recherche de Peukert avait observé que Max* inhibait l'interaction Miz-1/c-Myc. Il a été démontré qu'il existait une compétition entre Max* et Mid2s- β pour l'interaction avec c-Myc* par Bédard et al. en 2016. Dans la présente étude, il été montré que Mid2s- β semble compétitionner plus fortement pour le b-HLH-LZ de c-Myc que le b-HLH-LZ de Max, suggérant une très forte affinité. Nos présents résultats, supportés par le manuscrit de Bédard et collègues, viennent donc remettre en cause la théorie selon laquelle Miz-1, via son domaine Mid2 (acide aminés 641-713), formerait un complexe ternaire avec l'hétérodimère c-Myc/Max. Cependant, les résultats présentés viennent supporter la relation fonctionnelle entre c-Myc et Miz-1 établie dans la littérature et raffinent sa base structurale et mécanistique. De plus, les présents résultats font office de premiers pas dans l'élucidation d'une nouvelle approche pour le développement d'inhibiteurs de c-Myc ciblant son domaine b-HLH-LZ.

Comme expliqué en 1.3.2, le remplacement de résidus dans le domaine Mid2 possédant une forte tendance à former des structures hélicales (a.a. 695, 699, 702, 703, 706) avec une glycine et 4 prolines mène à l'abrogation de l'interaction en Miz-1 et le b-HLH-LZ de c-Myc. Cela suggère une interaction directe entre une partie en C-terminal du domaine Mid2 dans une conformation en hélice- α avec le HLH de c-Myc. Cependant, il a été démontré en 3.1.2 que la partie en C-terminal

du domaine Mid2 ne semble pas impliquée, mais plutôt la région centrale du Mid2. Une série de mutants de l'hélice 2 du HLH ont été démontrés comme altérant l'interaction Miz-1/c-Myc, plus précisément le mutant V394D. Comme mentionné en 3.1.5, le résidu V394 sur c-Myc* se situe dans la région d'agrégation prédite. Et comme la mutation V394D empêche la séquestration de c-Myc* du complexe c-Myc*/Max*/ADN (voir 3.1.3), il est possible d'affirmer que cette mutation retire la propensité de c-Myc* d'agréger en structures *cross-β* de type amyloïde avec Mid2s-β. Il est à noter que l'aspartate est un des pires résidus pour former des feuillets-β, ce qui vient supporter nos conclusions.

5. Conclusion

Au cours de ce mémoire, la compréhension de l'interaction entre le domaine Mid2 de Miz-1 et le bHLH-LZ de c-Myc a été raffinée. En isolant trois régions distinctes du domaine Mid2 selon ce qui était connu dans la littérature, il a été possible d'identifier une région minimale interagissant avec le bHLH-LZ de c-Myc : le peptide Mid2s-β. Ce peptide, englobant les résidus 660 à 692 du domaine Mid2, entraînait c-Myc* dans une conformation stable – ce que les deux autres peptides issus du Mid2 n'étaient capable d'effectuer. Nous avons alors montré que l'entraînement en complexe de c-Myc* par Mid2s-β permettait de perturber l'hétérodimère c-Myc*/Max*. Cela était dû au fait que c-Myc* possède une meilleure affinité pour Mid2s-β que pour Max*, et que le complexe c-Myc*/Mid2s-β possède une meilleure stabilité. De plus, l'hétérodimère c-Myc*/Max* en présence d'ADN se retrouve perturbé par le peptide Mid2s-β; sa présence séquestre c-Myc* dans un complexe ne pouvant lier l'ADN, tout en permettant à Max* d'homodimériser et de lier l'ADN. Dans la même série d'expérimentations, il a été démontré que la mutation V394D sur c-Myc* empêchait l'interaction avec Mid2s-β via l'abolition d'une surface hydrophobe et accessible au solvant. Finalement, il a été validé par RMN que le peptide Mid2s-β interagit spécifiquement avec c-Myc* sans affecter Max*, et que le complexe c-Myc*/Mid2s-β est de haut poids moléculaire, suggérant la formation de structures *cross-β* de type amyloïde (ce qui est appuyé par des résultats computationnels en 3.1.5).

La découverte d'une région minimale permettant de séquestrer c-Myc de complexes qui, en cellules humaines, activent la tumorigénèse constitue une avancée importante dans le développement d'inhibiteurs de c-Myc, puisque jusqu'à maintenant, l'oncogène est toujours considéré non-droguable dans la littérature. L'élucidation des déterminants structuraux inhérents au complexe c-Myc/Mid2s- β va permettre éventuellement de contourner les problèmes encourus par le développement de d'inhibiteurs de c-Myc par criblage à haut débit, et laisse entrevoir une nouvelle génération d'inhibiteurs permettant des thérapies ciblées contre le cancer.

5.1 Perspectives

Afin de mieux comprendre de quelle façon le peptide Mid2s- β séquestre le bHLH-LZ de c-Myc, il est nécessaire de bien comprendre les populations agrégées formées par le complexe c-Myc/Mid2s- β . Pour ce faire, il serait nécessaire d'effectuer des expérimentations en SAXS (voir 1.4.2) ou bien des expérimentations en DLS. Ces dernières seraient primordiales afin d'observer les espèces de haut poids moléculaire et d'établir une échelle de temps dynamique pour la formation de ces espèces. Ces résultats permettraient d'établir les conditions idéales afin d'obtenir des spectres RMN du complexe c-Myc*/Mid2- β de haute qualité, ce qui permettrait de calculer la structure 3D du complexe.

Liste des publications

Adams JM, Gerondakis S, Webb E, Corcoran LM, Cory S (1983) Cellular myc oncogene is altered by chromosome translocation to an immunoglobulin locus in murine plasmacytomas and is rearranged similarly in human Burkitt lymphomas. *Proc. Natl. Acad. Sci* 80: 1982-1986.

Adzhubei AA, Sternberg MJ (1993) Left-handed polyproline II helices commonly occur in globular proteins. *J Mol Biol* 229:472-493.

Adzhubei AA, Sternberg MJ (1994) Conservation of polyproline II helices in homologous proteins: implications for structure prediction by model building. *Protein Sci* 3:2395-2410.

Amati B, Littlewood TD, Evan GI, Land H (1993) The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. *EMBO Journal* 12(13): 5372-6.

Arvantis C, Felsher DW (2006) Conditional transgenic models define how MYC initiates and maintains tumorigenesis. *Seminars in Cancer Biology* 16: 313-317.

Aumiller WM, Keating CD (2016) Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nat. Chem* 8:129–137.

Bédard M, Maltais L, Montagne M, Lavigne P (2016) Miz-1 and Max Compete to Engage c-Myc: Implication for the Mechanism of Inhibition of c-Myc Transcriptional Activity by Miz-1. *Proteins* 85(2): 199-206.

Bender DA (2013) Le cancer : vue d'ensemble. *Biochimie de Harper*. De Boeck : Bruxelles (5^e éd), 725-747.

Berg T, Cohen SB, Desharnais J, Sonderegger C, Maslyar DJ, Goldberg J, Boger DL, Vogt PK (2002) Small-molecule antagonists of Myc/Max dimerization inhibit Myc-induced transformation of chicken embryo fibroblasts. *Proc Natl Acad Sci* 99(6):3830-5.

Bernard O, Cory S, Gerondakis S, Webb E, Adams JM (1983) Sequence of the murine and human cellular myc oncogenes and two modes of myc transcription resulting from chromosome translocation in B lymphoid tumours. *EMBO journal* 2(12): 2375-2383.

Bister K, Ramsay GM, Hayman MJ (1982) Deletions Within the Transformation-Specific RNA Sequence of Acute Leukemia Virus MC29 Give Rise to Partially Transformation-Defective Mutants. *Journal of Virology* 41: 754-766.

Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* 69:185-188.

Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, Tompa P, Fuxreiter M (2018) Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* 28:420–435.

Ramanathan A, Savol A, Burger V, Chenubothla CS, Agarwal PK (2014) Protein conformational populations and functionally relevant substates. *Acc Chem Res* 47(1):149-56.

Chandriani S, Frengen E, Cowling VH, Pendergrass SA, Perou CM, Whitfield ML, Cole MD (2009) A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. *PLoS One* 4(8): e6693.

Chauhan J, Wang H, Yap JL, Sabato PE, Hu A, Prochownik EV, Fletcher S (2014) Discovery of methyl 4'-methyl-5-(7-nitrobenzo[c][1,2,5]oxadiazol-4-yl)-[1,1'-biphenyl]-3-carboxylate, an improved small-molecule inhibitor of c-Myc-Max dimerization. *ChemMedChem* 9:2274-2285.
Chebrek R, Leonard S, Brevern AG, Gelly JC (2014) PolyprOnline: polyproline helix II and secondary structure assignment database. *Database* 2014:1-8

Chen J, Kriwacki RW (2018) Intrinsically Disordered Proteins: Structure, Function and Therapeutics. *Journal of Molecular Biology* 430: 2275-2277.

Chen R (2011) A General EMSA (Gel-shift) Protocol. *Bio-protocol* 1(3):e24.

Chevrier L, de Brevern A, Hernandez E, Leprince J, Vaudry H, Guedj AM, de Roux N (2013) PRR repeats in the intracellular domain of KISS1R are important for its export to cell membrane. *Mol Endocrinol* 27:1004-1014.

Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* 2007 8:65.

Cragnell C, Rieloff E, Skepö M (2018) Utilizing Coarse-Grained Modeling and Monte Carlo Simulations to Evaluate the Conformational Ensemble of Intrinsically Disordered Proteins and Regions. *J Mol Biol* 430(16):2478-2492.

Dang CV (2012) MYC on the Path to Cancer. *Cell* 149: 22-35.

Dalla-Favera R, Bregni M, Erikson J, Patterson D, Gallo RC, Croce CM (1982) Human c-myc oncogene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci. USA* 79:7824–7827.

D'Cruz CM, Gunther EJ, Boxer RB, Hartman JL, Sintasath L, Moody SE, Cox JD, Ha SI, Belka GK, Golant A, Cardiff RD, Chodosh LA (2001) c-MYC induces mammary tumorigenesis by means of a preferred pathway involving spontaneous Kras2 mutations. *Nature Medicine* 7(2): 235-239.

Dunker AK, Obradovic Z (2001) The protein trinity—linking function and disorder. *Nat Biotechnol* 19(9):805-806.

Eischen CM, Weber JD, Roussel MF, Sherr CJ, Cleveland JL (1999) Disruption of the Arf-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis. *Genes & Development* 13: 2658-2669.

Elam WA, Schrank TP, Campagnolo AJ, Hilser VJ (2013) Evolutionary conservation of the polyproline II conformation surrounding intrinsically disordered phosphorylation sites. *Protein Sci* 22: 405-417.

Enrietto PL, Hayman MJ (1982) Restriction Enzyme Analysis of Partially Transformation Defective Mutants of Acute Leukemia Virus MC29. *Journal of Virology* 44: 711-715.

Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki RW, Pappu RV, Brangwynne CP (2016) Coexisting liquid phases underlie nucleolar subcompartments. *Cell* 165: 1686–1697.

Ferré-D'Amaré AR, Prendergast GC, Ziff EB, Burley SK (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/LZ domain. *Nature* 363(6424):38-45.

Fuxreiter M (2018) Fuzziness in Protein Interactions – A Historical Perspective. *Journal of Molecular Biology* 430: 2278-2287.

Gomes E, Shorter J (2016) The molecular language of membraneless organelles. *J Biol Chem* 294(18):7115-7127.

Greenfield NJ (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols* 1(6):2527-2535.

Greenfield NJ (2007) Analysis of the kinetics of folding of proteins and peptides using circular dichroism. *Nature Protocols* 1(6):2891-2899.

Guo J, Li T, Schipper J, Nilson KA, Fordjour FK, Cooper JJ, Gordân R, Price DH (2014) Sequence specificity incompletely defines the genome-wide occupancy of Myc. *Genome Biology* 15(10): 482.

Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* 6:211-219.

Hamburger JB, Ferreon JC, Whitten ST, Hilser VJ (2004) Thermodynamic mechanism and consequences of the polyproline II (PII) structural bias in the denatured states of proteins. *Biochemistry*. 43(30):9790-9799.

Hanahan D, Weinberg RA (2000) The Hallmarks of Cancer. *Cell* 100(1): 57-70

Henriques J, Arleth L, Lindorff-Larsen K, Skepö M (2018) On the calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J Mol Biol* 430(16):2521-2539.

Herold S, Wanzel M, Beuger V, Frohme C, Beul D, Hillukkala T, Syvaioja J, Salus HP, Haenel F, Eilers M (2002) Negative Regulation of the Mammalian UV Response by Myc through Association with Miz-1. *Molecular Cell* 10: 509-521.

Holmstrom ED, Holla A, Zheng W, Nettels D, Best RB, Schuler B (2018) Accurate Transfer Efficiencies, Distance Distributions, and Ensembles of Unfolded and Intrinsically Disordered Proteins From Single-Molecule FRET. *Methods Enzymol* 611:287-325.

Hogg SJ, Newbold A, Vervoort SJ, Cluse LA, Martin BP, Gregory BP, Lefebure M, Vidacs E, Tothill RW, Bradner JE, Shortt J, Johnstone RW (2016) BET Inhibition Induces Apoptosis in Aggressive B-Cell Lymphoma via Epigenetic Regulation of BCL-2 Family Members. *Molecular Cancer Therapy* 15(9): 2030-41.

Horiuchi D, Anderton B, Goga A (2014) Taking on challenging targets: making MYC druggable. *Am Soc Clin Oncol Educ Book* e497-507.

Hyman AA, Weber CA, Jülicher F (2014) Liquid–liquid phase separation in biology. *Annu Rev Cell Dev Biol* 30:39–58.

Jacobsen NE (2007) *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*. John Wiley & Sons ISBN 978-0-471-73096-5.

Jain M, Arvanitis C, Chu K, Dewey W, Leonhardt E, Trinh M, Sundberg MD, Bishop JM, Felsher DW (2002) Sustained Loss of a Neoplastic Phenotype by Brief Inactivation of MYC. *Science* 297(5578): 102-104.

Janghorban M, Farrell AS, Allen-Petersen BL, Pelz C, Daniel CJ, Oddo J, Langer EM, Christensen DJ, Sears RC (2014) Targeting c-Myc by antagonizing PP2A inhibitors in breast cancer. *Proc. Natl. Acad. Sci* 111(25): 9157-62.

Ji H, Wu G, Zhan X, Nolan A, Koh C, De Marzo A, Doan HM, Fan J, Cheadle C, Fallahi M, Cleveland JL, Dang CV, Zeller KI (2011) Cell-type independent MYC target genes reveal a primordial signature involved in biomass accumulation. *PLoS One* 6(10): e26057.

Jung LA, Gebhardt A, Koelmel W, Ade CP, Walz S, Kuper J, von Eyss B, Letschert S, Redel C, d'Artista L, Biankin A, Zender L, Sauer M, Wolf E, Evan G, Kisker C, Eilers M (2016) OmoMYC blunts promoter invasion by oncogenic MYC to inhibit gene expression characteristic of MYC-dependent tumors. *Oncogene* 36:1911-1924.

Kalkat M, De Melo J, Hickman KA, Lourenco C, Redel C, Resette D, Tamachi A, Tu WB, Penn LZ (2017) Myc Dereglulation in Primary Human Cancers. *Genes* 8(6): 151.

Koh CM, Sabò A, Guccione E (2016) Targeting MYC in cancer therapy: RNA processing offers new opportunities. *Bioessays* 38(3):266-75.

Kohl NE, Kanda N, Schreck RR, Bruns G, Latt SA, Gilbert F, Alt FW (1983) Transposition and Amplification of Oncogene-Related Sequences in Human Neuroblastomas. *Cell* 35: 359-367.

Kosan C, Saba I, Godmann M, Herold S, Herkert B, Eilers M, Möröy T (2010) Transcription factor miz-1 is required to regulate interleukin-7 receptor signaling at early commitment stages of B cell differentiation. *Immunity* 33(6):917-28.

Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS Journal* 278:687-703.

Lavigne P, Kondejewski LH, Houston ME, Sönnichsen FD, Lix B, Sykes BD, Hodges RS, Kay CM (1995) Preferential Heterodimeric Parallel Coiled-coil Formation by Synthetic Max and c-Myc Leucine Zippers: A Description of Putative Electrostatic Interactions Responsible for the Specificity of Heterodimerization. *Journal of Molecular Biology* 254: 505-520.

Lavigne P, Crump MP, Gagné SM, Hodges RS, Kay CM, Sykes BD (1998) Insights into the mechanism of heterodimerization from the ¹H-NMR solution structure of the c-Myc-Max heterodimeric leucine zipper, *J. Mol. Biol.* 281:165–181.

Lavinia A. Carabet, Paul S. Rennie, Artem Cherkasov (2019) Therapeutics Inhibition of Myc in Cancer. *Structural Bases and Computer-Aided Drug Discovery Approaches. International Journal of Molecular Sciences* 20(1): 120.

Lin CY, Lovén J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA (2012) Transcriptional Amplification in Tumor Cells with Elevated c-Myc. *Cell* 151: 56-67

Luo J, Zhang C, Wang C, Li L, Li C, Li Q, Zhang M, Wu Q (2016) Miz-1 promotes the proliferation of esophageal cancer cells via suppression of p21 and release of p21-arrested cyclin D1. *Oncology Reports* 35: 3532-3540.

Maltais L, Montagne M, Bédard M, Tremblay C, Soucek L, Lavigne P (2017) Biophysical characterization of the b-HLH-LZ of ΔMax, an alternatively spliced isoform of Max found in tumor cells: Towards the validation of a tumor suppressor role for the Max homodimers. *PLoS ONE*, 12:e0174413.

Mitrea DM, and Kriwacki, RW (2016) Phase separation in biology; functional organization of a higher order. *Cell Commun Signal* 14:1.

Nair SK, Burley SK (2003) X-Ray Structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell*, 112:193-205.

Naud JF, Gagnon F, Wellinger R, Chabot B, Lavigne P (2003) Improving the Thermodynamic Stability of the Leucine Zipper of Max Increases the Stability of its b-HLH-LZ: E-box complex. *Journal of Molecular Biology* 326(5):1577-95.

Naud JF, McDuff FO, Sauvé S, Montagne M, Webb BA, Smith SP, Chabot B, Lavigne P (2005) Structural and Thermodynamical Characterization of the Complete p21 Gene Product of Max. *Biochemistry* 44(38): 12746-12758.

O'Shea EK, Klemm JD, Kim PS, Alber T (1991) X-Ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled-coil. *Science* 254(5031): 539-44.

Ozenne V, Schneider R, Yao M, Huang JR, Salmon L, Zweckstetter M, Jensen MR, Blackledge M (2012) Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *J Am Chem Soc* 134(36): 15138-48.

Paz A, Zeev-Ben-Mordehai T, Lundqvist M, Sherman E, Mylonas E, Weiner L (2008) Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3. *Biophysical Journal* 95 :1928–1944.

Peter S, Bultinck J, Myant K, Jaenicke LA, Walz S, Müller J, Gmachl M, Treu M, Boehmelt G, Ade CP, Schmitz W, Wiegner A, Otto C, Popov N, Sansom O, Kraut N, Eilers M (2014) Tumor cell-specific inhibition of MYC function using small molecule inhibitors of the HUWE1 ubiquitin ligase. *EMBO Mol Med* 6(12):1525-1541.

Peukert K, Staller P, Schneider A, Carmichael G, Hänel F, Eilers M (1997) An alternative pathway for gene regulation by Myc. *EMBO journal* 16(18): 5672-5686.

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. *Biophys J* 92:1439-1456.

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA (2010) c-Myc regulates transcriptional pause release. *Cell* 141(3): 432-45.

Ramsay GM, Enrietto PJ, Graf T, Hayman MJ (1982) Recovery of myc-specific sequences by a partially transformation-defective mutant of avian myelocytomatosis virus, MC29, correlates with the restoration of transforming activity. *Proc. Natl. Acad. Sci*, 79: 6885-6889.

Sauvé S, Tremblay L, Lavigne P (2004) The NMR Solution Structure of a Mutant of the Max b/HLH/LZ Free of DNA: Insights into the Specific and Reversible DNA Binding Mechanism of Dimeric Transcription Factors. *J Mol Biol* 342:813-832.

Sauvé S, Naud JF, Lavigne P (2007) The mechanism of Discrimination between Cognate and Non-Specific DNA by Dimeric b/HLH/LZ Transcription Factors. *Journal of Molecular Biology* 365: 1163-1175.

Seoane J, Pouponnot C, Staller P, Schader M, Eilers M, Massagué J (2001) TGFbeta influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b. *Nature Cell Biology* 3(4): 400-8.

Seoane J, Le HV, Massagué J (2002) Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature* 419(6908): 729-34.

Shi J, Stover JS, Whitby LR, Vogt PK, Bogger DL (2009) Small molecule inhibitors of Myc/Max dimerization and Myc-induced cell transformation. *Bioorg Med Chem Lett* 19(21):6038-41.

Shorter J, Lindquist S (2005) Prions as adaptive conduits of memory and inheritance. *Nat Rev Genet* 6:435-450.

Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* 274(5292):1531-1534

Staller P, Peukert K, Kiermaier A, Seoane J, Lukas J, Karsunky H, Möröy T, Bartek J, Massagué J, Hänel F, Eilers M (2001) Repression of p15^{INK4b} expression by Myc through association with Miz-1. *Nature Cell Biology* 3: 392-399.

Stogios PJ, Cuesta-Seijo JA, Chen L, Pomroy NC, Privé GG (2010) Insights into strand exchange in BTB domain dimers from the crystal structures of FAZF and Miz1. *J Mol Biol* 400(5):983-997

Stryer L, Berg JM, Tymoczko JL (2002) *Biochemistry*. WH Freeman, 5th edition, New York.

Taub R, Kirsch I, Morton C, Lenoir G, Swan D, Tronick S, Aaronson S, Leder P (1982) Translocation of the c-Myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proc. Natl. Acad. Sci.* 79: 7837-7841.

Tomasso ME, Tarver MJ, Devarajan D, Whitten ST (2016) Hydrodynamic Radii of Intrinsically Disordered Proteins Determined from Experimental Polyproline II Propensities. *PLoS Computational Biology* 12(1):e1004686.

Tompa M, Nagy A, Komoly S, Kalman B (2019) Wnt pathway markers in molecular subgroups of glioblastoma. *Brain Res* 1718:114-125.

Uversky VN, Gillespie JR, Fink AL (2000) Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41(3):415-27.

Uversky VN (2013) Disorder in the lifetime of a protein. *Intrinsically Disord Proteins* 1(1): e26782.

Uversky VN (2016) Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *The Journal of Biological Chemistry* 291(13): 6681-6688.

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005). The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* 59:687-696.

Wang H, Hammoudeh DI, Follis AV, Reese BE, Lazo JS, Metallo SJ, Prochownik EV (2007) Improved low molecular weight Myc-Max inhibitors. *Mol Cancer Therapy* 6:2399-2408.

Wang H, Chauhan J, Hu A, Pendleton K, Yap JL, Sabato PE, Jones JW, Perri M, Yu J, Cione E, Kane MA, Fletcher S, Prochownik EV (2013) Disruption of Myc-Max heterodimerization with improved cell-penetrating analogs of the small molecule 10074-G5. *Oncotarget* 4(6):936-47

Wanzel M, Herold S, Eilers M (2003) Transcriptional repression by Myc. *Trends Cell Biol* 13(3):146-50.

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337:635-645.

Weinstein IB (2002) Cancer. Addiction to oncogenes—the Achilles heel of cancer. *Science* 297(5578):63-4.

Weinstein IB, Joe A (2008) Oncogene Addiction. *Cancer Research* 68(9): 3077-3088.

Whitfield JR, Beaulieu ME, Soucek L (2017) Strategies to Inhibit Myc and Their Clinical Applicability. *Front Cell Dev Biol* 5:10.

Whitmore L, Wallace BA (2008) Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and References Databases. *Biopolymers* 89:392-400.

Wiese KE, Walz S, Eyss BJ, Wolf E, Athineos D, Sansom O, Eilers M (2013) The Role MIZ-1 in MYC-Dependent Tumorigenesis. *Cold Spring Harbor Perspective in Medecine* 3:a014290.

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* 73:1-16.

Wolf E, Gebhardt A, Kawauchi D, Walz S, von Eyss B, Wagner N, Renninger C, Krohne G, Asan E, Roussel MF, Eilers M (2013) Miz-1 is required to maintain autophagic flux. *Nat Comm* 4:2535.

Woody RW (2009) Circular dichroism spectrum of peptides in the poly(Pro)II conformation. *J. Am. Chem. Soc.* 131: 8234–8245.

Yap JL, Wang H, Hu A, Chauhan J, Jung KY, Gharavi RB, Prochownik EV, Fletcher S (2013) Pharmacophore identification of c-Myc inhibitor 10074-G5. *Bioorgan Med Chem Lett* 23:370-374.

Yeh E, Cunningham M, Arnold H, Chasse D, Monteith T, Ivaldi G, Hahn WC, Stukenberg PT, Shenolikar S, Uchida T, Counter CM, Nevins JR, Means AR, Sears R (2004) A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nature Cell Biology* 6(4): 308-18.

Yu C, Niu X, Jin F, Liu Z, Jin C, Lai L (2016) Structure-based Inhibitor Design for the Intrinsically Disordered Protein c-Myc. *Sci Rep* 6:22298.

Zheng W, Hofmann H, Schuler B, Best RB (2018) Origin of Internal Friction in Disordered Proteins Depends on Solvent Quality. *J Phys Chem B* 122(49):11478-11487.

Ziarek JJ, Peterson FC, Lytle BL, Volkman BF (2011) Binding site identification and structure determination of protein-ligand complexes by NMR. *Methods Enzymol* 493:241-275.